

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524d-30

Perfect score: 79

Sequence: 1 MNIKLKMPLYAGYK 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	374	2 JC2124	major allergen Cry
2	79	100.0	374	2 JC2123	major allergen Cry
3	44	55.7	1202	2 T23429	hypothetical prote
4	43	54.4	345	2 T28026	hypothetical prote
5	41	51.9	252	2 H26965	short-chain dehydr
6	40	50.6	522	2 H86248	protein T23U18.22
7	39	49.4	67	2 S20967	hypothetical prote
8	39	49.4	162	2 B18177	probable pili NMA
9	39	49.4	294	2 H38888	COI intron 15 prot
10	39	49.4	347	2 S59837	probable membrane
11	39	49.4	395	2 AH0441	probable exported
12	39	49.4	528	2 UC1118	alkyl sulfatase (E
13	39	49.4	757	2 T16609	hypothetical prote
14	38	48.1	149	2 S48927	hypothetical prote
15	38	48.1	268	2 A83651	hypothetical prote
16	38	48.1	320	2 P86821	hypothetical prote
17	38	48.1	541	2 DB2302	hypothetical prote
18	38	48.1	614	1 T26303	iron(III) ABC tran
19	38	48.1	646	1 W2BEC8	probable membrane
20	38	48.1	647	2 T42579	68.6K capsid prote
21	37	46.8	157	2 H87601	transcription elon
22	37	46.8	166	1 KRB02B	keratin, 68K type
23	37	46.8	167	1 S21359	keratin, type I, c
24	37	46.8	182	1 KRB02A	keratin, type I, c
25	37	46.8	182	2 A02946	keratin, 59K type
26	37	46.8	300	2 AC2235	hypothetical prote
27	37	46.8	338	2 G69413	conserved hypochet
28	37	46.8	348	2 C96673	gamma-tocopherol m
29	37	46.8	363	2 B97199	uncharacterized co

30	37	46.8	370	2 A25004	keratin, 53K type
31	37	46.8	384	2 I61769	keratin ed, type I
32	37	46.8	390	2 A60093	cytochrome, type
33	37	46.8	400	2 AD3293	opgc protein limpo
34	37	46.8	475	2 T45766	hypothetical prote
35	37	46.8	483	2 A34720	keratin 8, type II
36	37	46.8	493	2 H72165	hypothetical prote
37	37	46.8	493	2 A36850	A18R protein - var
38	37	46.8	493	2 T28560	hypothetical prote
39	37	46.8	502	2 A23547	keratin, type II c
40	37	46.8	511	2 A39340	neurofilament prot
41	37	46.8	513	2 S08381	keratin, 58K type
42	37	46.8	520	2 JS0291	intermediate filam
43	37	46.8	524	2 A23518	keratin, 57K type
44	37	46.8	553	2 I59009	epidermal keratin
45	37	46.8	555	2 T20137	hypothetical prote

## ALIGNMENTS

### RESULT 1

JC2124 major allergen Cry j I precursor (clone pCCI-15) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_reviseion 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2124

R/Sone, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan

A/Reference number: JC2123; PMID:94183234; PMID:8135802

A/Accession: JC2124

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:D26545; NID:9493633; PIDN:BA05543.1; PID:9493634

A/Experimental source: pollen

A/Note: the authors described carbohydrate binding site for residue 279

C/Suprafamily: pectate lyase IAT59

C/Keywords: glycoprotein; pollen

F.1-21/Domain: signal sequence #status predicted <SIG>

F.22-374/Product: major allergen Cry j I (clone pCCI-15) #status predicted <MAT>

F.158-191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 79; DB 2; Length 374;

Best local similarity 100.0%; Pred. No. 9.7e-07; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPLYAGYK 15

DB 97 MNIKLKMPLYAGYK 111

RESULT 2

JC2123 major allergen Cry j I precursor (clone pCCI-2-2) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_reviseion 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2123; PMID:94183234; PMID:8135802

R/Sone, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan

A/Reference number: JC2123; PMID:94183234; PMID:8135802

A/Accession: JC2123

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:D26544; NID:9493631; PIDN:BA05542.1; PID:9493632

A/Experimental source: pollen

A/Note: the authors described carbohydrate binding site for residue 279

C/Suprafamily: pectate lyase IAT59

C/Keywords: glycoprotein; pollen

F:1-21/Domain: signal sequence #status predicted <STIG>  
 F:22-37/Product: major allergen Cry I (clone pCCT-2-2) #status predicted <MAN>  
 F:158,151,253,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 79; DB 2; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NWIKKMPYIAGYK 15  
 DB 97 NWIKKMPYIAGYK 111

## RESULT 3

hypothetical protein K07G5.3 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C/Accession: T23429  
 R:McMurray, A.  
 submitted to the EMBL Data Library, April 1996  
 A/Reference number: Z19739  
 A/Accession: T23429  
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
 A/Residues: 1-1202 <WTL>  
 A/Cross-references: EMBL:Z71264; PIDN:CAA95830.1; GSPDB:GN00019; CESP:K07G5.3  
 A/Experimental source: clone K07G5  
 C/Genetics:  
 A/Gene: CESP:K07G5.3  
 A/Map position: 1  
 A/Introns: 13/1; 104/3; 196/3; 260/2; 282/3; 307/2; 361/2; 410/2; 535/2; 603/2; 688/1; 7

C/Superfamily: Caenorhabditis elegans hypothetical protein K07G5.3

Query Match 55.7%; Score 44; DB 2; Length 1202;  
 Best Local Similarity 53.8%; Pred. No. 11;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 IKLKMPYIAGYK 15  
 DB 811 IRLNIPLYISNYK 823

## RESULT 4

hypothetical protein ZK829.8 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C/Accession: T28026  
 R:Harris, B.  
 submitted to the EMBL Data Library, May 1996  
 A/Reference number: Z20458  
 A/Accession: T28026  
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
 A/Residues: 1-345 <WTL>  
 A/Cross-references: EMBL:Z73899; PIDN:CAA98078.1; GSPDB:GN00022; CESP:ZK829.8  
 A/Experimental source: clone ZK829  
 C/Genetics:  
 A/Gene: CESP:ZK829.8  
 A/Map position: 4  
 A/Introns: 64/3; 120/2; 188/3; 277/2

Query Match 54.4%; Score 43; DB 2; Length 345;  
 Best Local Similarity 53.8%; Pred. No. 4.3;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 IKLKMPYIAGYK 15  
 DB 30 LKFKSPYIIGYR 42

## RESULT 5

H96965

short-chain dehydrogenase (gene dltE) [imported] - Clostridium acetobutylicum  
 C/Species: Clostridium acetobutylicum  
 C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C/Accession: H96965  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Le  
 J. Bacteriol. 183, 4823-4838, 2001

A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
 A/Reference number: A96900; MUID:21359325; PMID:21359325  
 A/Accession: H96965  
 A/Status: preliminary

A/Molecule type: DNA  
 A/Residues: 1-252 <KUR>  
 A/Cross-references: GB:AE001437; PIDN:AAK78515.1; PID:G15023401; GSPDB:GN00168  
 A/Experimental source: Clostridium acetobutylicum ATCC824  
 C/Genetics:  
 A/Gene: CAC0536

Query Match 51.9%; Score 41; DB 2; Length 252;  
 Best Local Similarity 53.8%; Pred. No. 7.3;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NWIKKMPYIAG 13  
 DB 109 VDLNKKAPYIAG 121

## RESULT 6

H86248  
 protein T23J18.22 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: H86248  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, T.  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Malt, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Kowley, D.; Sakano, H.  
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712  
 A/Accession: H86248  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-522 <STO>  
 A/Cross-references: GB:AE005172; NID:96554189; PIDN:AAF16635.1; GSPDB:GN00141  
 C/Genetics:  
 A/Gene: T23J18.22  
 A/Map position: 1

Query Match 50.6%; Score 40; DB 2; Length 522;  
 Best Local Similarity 57.1%; Pred. No. 24;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NIKLKMPYIAGYK 15  
 DB 240 NIREMPYIKDYE 253

## RESULT 7

S20967  
 hypothetical protein 3 - Chlamydomonas reinhardtii chloroplast  
 C/Species: chloroplast Chlamydomonas reinhardtii  
 C/Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 18-Jun-1993

C/Accession: S20967  
 R:Choquet, Y.; Rahire, M.; Girard-Bascou, J.; Erickson, J.; Rochaix, J.D.  
 EMBO J. 11, 1697-1704, 1992  
 A/Title: A chloroplast gene is required for the light-independent accumulation of chlorophyll  
 A/Reference number: S20965; MUID:92258378; PMID:1374710

A:Accession: S20967  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-67 <CHO>  
 C:Genetics:  
 A:Genome: chloroplast  
 C:Keywords: chloroplast

Query Match 49.4%; Score 39; DB 2; Length 67;  
 Best Local Similarity 41.7%; Pred. No. 4.1;  
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Db 12 NWKMLPYAGK 13  
 12 NWKMLPYAGK 23

RESULT 8  
 B81877  
 Probable piliN NMA1110 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)  
 C:Species: Neisseria meningitidis  
 C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: B81877  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
 ; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A:Reference number: A81775; MUID:20222556; PMID:10761919  
 A:Accession: B81877  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-162 <PAR>  
 A:Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CAB84372.1; PID:G737980  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: NMA1110

Query Match 49.4%; Score 39; DB 2; Length 162;  
 Best Local Similarity 46.2%; Pred. No. 11;  
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Db 72 IKSLKMFVSGK 84  
 72 IKSLKMFVSGK 84

RESULT 9  
 H38888  
 COI intron 15 protein - Podospora anserina mitochondrion  
 C:Species: mitochondrion Podospora anserina  
 C>Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 21-Jul-2000  
 C:Accession: H38888  
 R:Cummings, D.J.; Michel, F.; McNally, K.L.  
 Curr. Genet. 16, 381-406, 1989  
 A>Title: DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I mt  
 A:Reference number: A48327; MUID:90124722; PMID:2558609  
 A:Accession: H38888  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-294 <CUM>  
 A:Cross-references: GB:X55026; GB:M30937; GB:M61734; NID:G14030; PIDN:CMA38793.1; PID:G1  
 A:Genetics:  
 A:Genome: mitochondrion  
 A:Genetic code: SGC3  
 C:Keywords: mitochondrion

Query Match 49.4%; Score 39; DB 2; Length 294;  
 Best Local Similarity 53.8%; Pred. No. 20;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Db 191 NIDKIPYFKGW 203  
 191 NIDKIPYFKGW 203

## RESULT 10

S59837  
 Probable membrane protein YPR180W - Yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein P9705.5  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 06-Feb-1998  
 C:Accession: S59837  
 R:Pauley, A.  
 submitted to the EMBL Data Library, April 1995  
 A:Description: The sequence of S. cerevisiae cosmid 9705.  
 A:Reference number: S59829  
 A:Accession: S59837  
 A:Molecule type: DNA  
 A:Residues: 1-347 <PAU>  
 A:Cross-references: EMBL:U25842; NID:G786312; PID:G786317; MIPS:YPR180W  
 C:Genetics:  
 A:Gene: SGD:AOS1  
 A:Cross-references: SGD:S0006384; MIPS:YPR180W  
 A:Map position: 16R  
 C:Keywords: transmembrane protein  
 F:306-322/Domain: transmembrane #status predicted <TM>

Query Match 49.4%; Score 39; DB 2; Length 347;  
 Best Local Similarity 60.0%; Pred. No. 24;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 145 KLNIPLYVAG 154  
 145 KLNIPLYVAG 154

## RESULT 11

AH0441  
 Probable exported protein YPO3631 [imported] - Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C:Accession: AH0441  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.  
 demo-tariga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
 Nature 413, 523-527, 2001  
 A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001, MUID:21470413; PMID:11586360  
 A:Accession: AH0441  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-395 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC93100.1; PID:G15981552; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: YPO3631

Query Match 49.4%; Score 39; DB 2; Length 395;  
 Best Local Similarity 54.5%; Pred. No. 28;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 320 LRLPYIAGFE 330  
 320 LRLPYIAGFE 330

## RESULT 12

JC1118  
 alkyl sulfatase (EC 3.1.6.-) - Pseudomonas sp.  
 N:Alternate names: SDSase, sdsA protein  
 C:Species: Pseudomonas sp.  
 C>Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 08-Oct-1999  
 C:Accession: JC1118; S27644  
 R:Davidson, J.; Brunel, F.; Phanopoulos, A.; Prozzi, D.; Terpestra, P.  
 Gene 114, 19-24, 1992  
 A>Title: Cloning and sequencing of Pseudomonas genes determining sodium dodecyl sulfate  
 A:Reference number: JC1118; MUID:92267380; PMID:1587481

A:Accession: JCI118  
 A:Molecule type: DNA  
 A:Residues: 1-528 <DAV>  
 A:Cross-references: GB:M86744; NID:g151550; PIDN:AA25989.1; PID:g151552  
 A:Experimental source: ATCC 19151  
 C:Genetics:  
 A:Gene: sdsA  
 C:Keywords: sulfuric ester hydrolase

Query Match 49.4%; Score 39; DB 2; Length 528;  
 Best Local Similarity 54.5%; Pred. No. 38;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 4 KTKMPYIAGY 14  
 DB 257 KVKLPYIAGY 267

# RESULT 13

TI6609  
 hypothetical protein K10B3.6 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
 C:Accession: T16609  
 R:Gatung, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Description: The sequence of C. elegans cosmid K10B3.  
 A:Reference number: Z18546  
 A:Accession: T16609  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-757 <GAT>  
 A:Cross-references: EMBL:U49941; NID:g1206038; PID:g1206045; PIDN:AAB53873.1; GSPDB:GN00  
 A:Experimental source: strain Bristol N2; clone K10B3  
 C:Genetics:  
 A:Gene: CESP.K10B3.6  
 A:Map position: X  
 A:Insertions: 89/2; 142/2; 184/2; 217/1; 269/3; 292/3; 330/3; 377/1; 425/2; 456/3; 497/3; 5

Query Match 49.4%; Score 39; DB 2; Length 757;  
 Best Local Similarity 54.5%; Pred. No. 56;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 NTKIKMPYI 11  
 DB 558 INIKYPMY 568

# RESULT 14

S48927  
 hypothetical protein YHL041w - yeast (*Saccharomyces cerevisiae*)  
 C:Species: *Saccharomyces cerevisiae*  
 C>Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 19-Apr-2002  
 C:Accession: S48927  
 R:Favella, T.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: The sequence of S. cerevisiae cosmid 9196.  
 A:Reference number: S46794  
 A:Accession: S48927  
 A:Molecule type: DNA  
 A:Residues: 1-149 <FAV>  
 A:Cross-references: EMBL:U11583; NID:g2289854; PID:g2289870; GSPDB:GN00008; MIPS:YHL041w  
 C:Genetics:  
 A:Gene: MIPS.YHL041w  
 A:Cross-references: SGD:S0001033  
 A:Map position: 8L  
 C:Superfamily: *Saccharomyces* hypothetical protein YHL041w

Query Match 48.1%; Score 38; DB 2; Length 149;  
 Best Local Similarity 54.5%; Pred. No. 15;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 4 KTKMPYIAGY 14

DB 55 KIKTPYIAGY 65

# RESULT 15

A83651  
 hypothetical protein BH0009 [imported] - *Bacillus halodurans* (strain C-125)  
 C:Species: *Bacillus halodurans*  
 C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: A83651  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: A83651  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-268 <STO>  
 A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA803728.1; GSPDB:GN  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0009

Query Match 48.1%; Score 38; DB 2; Length 268;  
 Best Local Similarity 53.8%; Pred. No. 28;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 NTKIKMPYIAGY 14  
 DB 162 SIQKBPVAVAGY 174

Search completed: April 20, 2003, 13:15:42  
 Job time: 8.0785 secs



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein & protein search, using sw model

Run on: April 20, 2003, 12:54:25; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524D-30  
Perfect score: 79  
Sequence: 1 MNILKMPMTAGYK 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	374	1 SBP_CRYUA	P18632 cryptomeria
2	71	89.9	346	1 MPAL_CUPAR	Q9639 cupressus a
3	70	88.6	367	1 MPAL_JUNAS	P81234 juniperus a
4	59	74.7	375	1 MPAL_CHAOB	Q96385 chamaecypar
5	40	50.6	627	1 K2C1_MOUSE	P04104 mus musculu
6	39	49.4	347	1 RH31_YEAST	Q06624 saccharomyc
7	38	48.1	149	1 YHE1_YEAST	P38730 caenorhabdi
8	38	48.1	614	1 YHE2_YEAST	Q09320 saccharomyc
9	38	48.1	614	1 YHE2_YEAST	P28936 equine herp
10	38	48.1	166	1 K2C5_BOVIN	P04262 bos taurus
11	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
12	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
13	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
14	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
15	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
16	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
17	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
18	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
19	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
20	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
21	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
22	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
23	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
24	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
25	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
26	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
27	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
28	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
29	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
30	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
31	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
32	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus
33	37	46.8	182	1 K2C4_BOVIN	P04263 bos taurus

34	37	46.8	638	1 K220_HUMAN	Q01546 homo sapien
35	37	46.8	645	1 K22E_HUMAN	P35908 homo sapien
36	37	46.8	1203	1 Y741_CAEEL	O11069 caenorhabdi
37	37	46.8	2214	1 POLG_CXAZ4	P36290 c genome po
38	36	45.6	188	1 EFP_RICPR	Q92677 rickettsia
39	36	45.6	273	1 YFCO_ECOLI	P76498 escherichia
40	36	45.6	358	1 Y993_METUA	Q58400 methanococc
41	36	45.6	389	1 TPEA_SULSH	Q05208 sulfolobus
42	36	45.6	422	1 K1CR_MOUSE	P05783 mus musculu
43	36	45.6	429	1 K1CR_MOUSE	P23239 xenopus lae
44	36	45.6	458	1 DESM_XENLA	P57002 neisseria m
45	36	45.6	460	1 PGMU_XENLA	

## ALIGNMENTS

RESULT 1  
ID SBP\_CRYUA STANDARD, PRT, 374 AA.

AC P18632, 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Sugi basic protein precursor (SBP) (Major allergen Cry j 1) (Cry j I).  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
ON NCBI\_TaxID=3369;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Pollen;  
RX MEDLINE=94183234; PubMed=8135802;  
RA Some T., Komiyama N., Shimizu K., Kusakabe T., Morikubo K.,  
RT Kuno K.,  
RT "Cloning and sequencing of cDNA coding for Cry j I, a major allergen  
of Japanese cedar pollen.",  
RL Biochem. Biophys. Res. Commun. 199;619-625(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pollen;  
RX MEDLINE=95003748; PubMed=7920021;  
RA Hijioka A., Matsumoto I., Kojima K., Ogawa H.,  
RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
(Cryptomeria japonica) pollen allergen, Cry j I.",  
RL Int. Arch. Allergy Immunol. 105;198-202(1994).  
RN [5]  
RP STRUCTURE OF CARBOHYDRATES.  
RC TISSUE=Pollen;  
RX MEDLINE=95332249; PubMed=7608114;  
RA Hino K., Yamamoto S., Sano O., Taniguchi Y., Kohno K., Usui M.,  
RT "Carbohydrate structures of the glycoprotein allergen Cry j I from  
Japanese cedar (Cryptomeria japonica) pollen.",  
RL J. Biochem. 117;289-295(1995).  
CC -1- P1M: CONTAINS FUCCOSE/XULOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
CC -1- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR  
CC POLLEN. THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
CC -1- MISCELLANEOUS: THE SEQUENCE OF CRY J I FORM A IS SHOWN HERE. FORM

B DIFFERS IN SIX POSITIONS.  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D26544; BAA05542.1; -  
CC EMBL; D26545; BAA05543.1; -  
CC EMBL; D34639; BAA07020.1; -  
CC PIR; A44773; A44773.  
CC DR GlycoSiteDB; P18632; -  
CC DR InterPro; IPR002022; Amb\_allergen.  
CC DR Pfam; PF00544; pec\_lyase; 1.  
CC DR PRINTS; PR00807; AMBALLERGEN.  
CC Allergen; Glycoprotein; Multigene family; signal.  
KW SIGNAL  
FT 1 21  
FT CHAIN 22 374  
FT CARBOHYD 158 158  
FT CARBOHYD 191 191  
FT CARBOHYD 293 293  
FT CARBOHYD 354 354  
FT VARIANT 12 12  
FT VARIANT 143 143  
FT VARIANT 202 202  
FT VARIANT 221 221  
FT VARIANT 358 358  
FT VARIANT 361 361  
SQ SEQUENCE 374 AA; 40645 MW; 74AB25950248F56F CRC64;  
Query Match 100.0%; Score 79; DB 1; Length 374;  
Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNITKMPYIAGYK 15  
DB 97 MNITKMPYIAGYK 111  
RESULT 2  
MPAL\_CUPAR STANDARD; PRT; 346 AA.  
AC Q9SCG9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Major pollen allergen Cup a 1.  
OS Cupressus arizonica.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferales; Cupressaceae; Cupressus.  
OX NCBI\_TaxId=99011;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20571526; PubMed=1122214;  
RA Aceituno E., Del Pozo V., Minguez A., Arieta I., Cortegano I.,  
RA Cardaba B., Gallardo S., Rojo M., Palomino P., Lahoz C.;  
RT Molecular cloning of major allergen from Cupressus arizonica pollen;  
RT Cup a 1.";  
RT Clin. Exp. Allergy 30:1750-1758(2000).  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AJ243570; CAB62551.1; -  
CC InterPro; IPR002022; Amb\_allergen.  
CC Pfam; PF00544; pec\_lyase; 1.  
CC DR PRINTS; PR00807; AMBALLERGEN.  
KW Allergen; Glycoprotein.  
FT CARBOHYD 127 127  
FT CARBOHYD 157 157  
FT CARBOHYD 272 272  
SQ SEQUENCE 346 AA; 37589 MW; F1281DDDA1D5DFD0 CRC64;  
Query Match 89.9%; Score 71; DB 1; Length 346;  
Best Local Similarity 80.0%; Pred. No. 6.3e-06;  
Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNITKMPYIAGYK 15  
DB 76 MNITKMPYIAGYK 90  
RESULT 3  
MPAL\_JUNAS STANDARD; PRT; 367 AA.  
AC P81254; Q9ZNU7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Major pollen allergen Jun a 1 precursor.  
OS Juniperus ashei (Ozark white cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferales; Cupressaceae; Juniperus.  
OX NCBI\_TaxId=11101;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-38; 42-50; 58-80; 88-94;  
RP 117-124; 134-140; 160-164; 256-263 AND 322-325.  
RX MEDLINE=99414163; PubMed=10482836;  
RA Midoro-Horiuchi T.M., Goldblum R.M., Kurosky A., Wood T.G.,  
RA Brooks E.G.;  
RT Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
RT allergen, Jun a 1.";  
RT J. Allergy Clin. Immunol. 104:608-612(1999).  
CC -1- DISASS: THIS PROTEIN CAUSES SEVERAL SEASONAL ALLERGIC RHINITIS  
CC IN NORTH AMERICA.  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF106663; AAD03609.1; -  
CC EMBL; AF106662; AAD03608.1; -  
CC InterPro; IPR002022; Amb\_allergen.  
CC Pfam; PF00544; pec\_lyase; 1.  
CC DR PRINTS; PR00807; AMBALLERGEN.  
KW Allergen; Glycoprotein; signal.

FT SIGNAL 1 21  
 FT CHAIN 22 367 MAJOR POLLEN ALLERGEN JUN A 1.  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 367 AA; FC9B81E675662E49 CRC64;  
 Query Match  
 Best Local Similarity 88.6%; Score 70; DB 1; Length 367;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 NNKIKMPYIAGYK 15  
 Db 97 NNKIKMPYIAGYK 111

RESULT 4  
 MPAL CHAOB STANDARD; PRT; 375 AA.  
 ID MPAL CHAOB  
 AC 096385;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major pollen allergen Cha o 1 precursor.  
 OS Chamaecyparis obtusa (Japanese cypress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae;  
 OC Chamaecyparitis.  
 OC NCBI\_TaxID=13415;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Pollen;  
 RA MEDLINE=96265194; Pubmed=8676896;  
 RA Suzuki M., Komiyama N., Itoh H., Sone T., Kuno K., Takagi I.,  
 RA Ohta N.,  
 RT "Purification, characterization and molecular cloning of Cha o 1, a  
 RT major allergen of Chamaecyparis obtusa (Japanese cypress) pollen.";  
 RL Mol. Immunol. 33:451-460(1996).  
 CC -1- SIMILARITY: BELONGS TO THE POLYACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Da5404; BAA08246.1; -  
 CC InterPro: IPR002022; Amb allergen.  
 CC Pfam: PF00544; pec lyase 1.  
 CC PRINTS: PRO0807; AMBALLERGEN.  
 CC Allergen: Glycoprotein; Signal.  
 CC  
 CC FT CHAIN 1 21  
 CC FT CARBOHYD 22 375 MAJOR POLLEN ALLERGEN CHA O 1.  
 CC FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC SQ SEQUENCE 375 AA; 40258 MW; 81CD91DF7066DBF CRC64;

Query Match  
 Best Local Similarity 74.7%; Score 59; DB 1; Length 375;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 NNKIKMPYIAGYK 15  
 Db 97 NNKIKMPYIAGYK 111

RESULT 5  
 K2C1\_MOUSE

ID K2C1\_MOUSE STANDARD; PRT; 627 AA.  
 AC P04104;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Keratin, type II cytoskeletal I (CytoKeratin I) (67 kDa cytoKeratin).  
 GN KRT1 OR KRT2-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE OF 83-628 FROM N.A.  
 RX MEDLINE=85207740; Pubmed=2581964;  
 RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,  
 RA Roop D.R.;  
 RT "Amino acid sequences of mouse and human epidermal type II keratins  
 RT of Mr 67,000 provide a systematic basis for the structural and  
 RT functional diversity of the end domains of Keratin Intermediate  
 RT filament subunits.";  
 RL J. Biol. Chem. 260:7142-7149(1985).  
 RN [2]  
 RP REVISIONS, AND SEQUENCE FROM N.A.  
 RA Roop D.R.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: HETEROHEPTAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
 CC MICROFILAMENTAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)  
 CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M10937; AAD05191.1; -  
 CC PIR: A02951; KMS2.  
 CC SWISS-2DPAGE: P04104; MOUSE.  
 CC MGD: MGI:96698; Krt2-1.  
 CC InterPro: IPR001664; IF.  
 CC Pfam: PF00038; filament; 1.  
 CC PROSITE: PS00226; IF; 1.  
 CC KW Intermediate filament; Coiled coil; Keratin.  
 CC FT INIT MET 0  
 CC FT DOMAIN 1 186 HEAD.  
 CC FT DOMAIN 187 496 ROD.  
 CC FT DOMAIN 497 627 TAIL.  
 CC FT DOMAIN 187 222 COIL 1A.  
 CC FT DOMAIN 223 242 LINKER 1.  
 CC FT DOMAIN 243 333 COIL 1B.  
 CC FT DOMAIN 334 357 LINKER 12.  
 CC FT DOMAIN 358 496 COIL 2.  
 CC FT SITE 451 451 STUTTER.  
 CC SQ SEQUENCE 627 AA; 65092 MW; E7B848654539578 CRC64;  
 Query Match  
 Best Local Similarity 50.6%; Score 40; DB 1; Length 627;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 NNKIKMPYIAGYK 15  
 Db 476 NNKIKMPYIAGYK 490

RESULT 6  
 RH31\_YEAST STANDARD; PRT; 347 AA.  
 ID RH31\_YEAST  
 AC Q06624;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA damage tolerance protein RHC31 (RAD31 homolog).  
 GN AOS1 OR RHC31 OR YPR180W OR P9705.5.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
 NCBI\_TaxID=4932;  
 RX MEDLINE=97333271; PubMed=9169875;  
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,  
 RA Basseg H., Aparicio A., Barrall B.G., Badcock K., Barnes V.,  
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
 RA Dietrich F.S., Deltis H., DiPaolo J., Dubois E., Duetschhoff A.,  
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,  
 RA Hunkle-Smith S., Hyman R., Johnston M., Kallman S., Kleene K.,  
 RA Komp C., Kurdi O., Laekari D., Lew H., Lin A., Lin D., Louis E.J.,  
 RA Matthe R., Messerguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
 RA Meiller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
 RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,  
 RA Scherens B., Schramm S., Schroeder W., Sidou A.M., Tettelin H.,  
 RA Urestrazaru L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
 RA Walsh S.W., Wambutt R., Wang Y., Wedler H., Wilmott E.,  
 RA Zhong W.V., Zollner A., Vo D.H., Han J.;  
 RA "the nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";  
 RL Nature 387:103-105(1997).  
 RX MEDLINE=97248684; PubMed=9092625;  
 RA Shayed M., Does C.L., Tavares M., Watts F.Z.;  
 RT "Characterisation of Schizosaccharomyces pombe rad31, a UBA-related  
 RT gene required for DNA damage tolerance.";  
 RL Nucleic Acids Res. 25:1162-1169(1997).  
 CC -1- FUNCTION: COULD BE INVOLVED IN A UBIQUITIN-RELATED PROCESS  
 CC -1- IMPORTANT FOR DNA DAMAGE TOLERANCE.  
 CC -1- SIMILARITY: TO THE N-TERMINAL OF UBIQUITIN-ACTIVATING ENZYME E1.  
 CC -1- SIMILARITY: STRONG, TO S.POMBE RAD31.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U25842; AAB68113.1; -  
 DR SGI; S0006384; AOS1.  
 DR InterPro; IPR000594; Thif\_domain.  
 DR Pfam; PF00899; Thif; 1.  
 KW DNA damage.  
 SQ SEQUENCE 347 AA; 39273 MW; 1ADB7B817BA27F3E CRC64;  
 QY 4 KLMMPYIAG 13  
 DB 145 KLMPIYVAG 154  
 ID YH01\_YEAST STANDARD; PRT; 149 AA.  
 AC P38730;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE Hypothetical 17.6 kDa protein in CBP2 5' region.  
 GN YH041W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
 NCBI\_TaxID=4932;  
 RX MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Lacroix P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
 RA Nham M., Rifkin L., Riley L., St Peter H., Trevaskis E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohlman P., Waterson R., Wilson R.,  
 RA Vaudin M.;  
 RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT VIII.";  
 RL Science 265:2077-2082(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U11583; AAB65053.1; -  
 DR PIR; S48927; S48927.  
 DR SGI; S0001033; YH041W.  
 KW Hypothetical protein.  
 SQ SEQUENCE 149 AA; 17554 MW; 5DE657D38A83CFDF CRC64;  
 QY 4 KLMMPYIAG 14  
 DB 55 KLMPIYVAG 65  
 ID Y022\_CAEEL STANDARD; PRT; 614 AA.  
 AC Q09320;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 69.0 kDa protein F40B5.2 in chromosome X.  
 GN F40B5.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditidae.  
 NCBI\_TaxID=6239;  
 RX MEDLINE=97333271; PubMed=9169875;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Lacroix P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
 RA Nham M., Rifkin L., Riley L., St Peter H., Trevaskis E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohlman P., Waterson R., Wilson R.,  
 RA Vaudin M.;  
 RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT VIII.";  
 RL Science 265:2077-2082(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U25842; AAB68113.1; -  
 DR SGI; S0006384; AOS1.  
 DR InterPro; IPR000594; Thif\_domain.  
 DR Pfam; PF00899; Thif; 1.  
 KW DNA damage.  
 SQ SEQUENCE 347 AA; 39273 MW; 1ADB7B817BA27F3E CRC64;  
 QY 4 KLMMPYIAG 13  
 DB 145 KLMPIYVAG 154  
 ID YH01\_YEAST STANDARD; PRT; 149 AA.  
 AC P38730;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)

DR EMBL: U23182; AAA64337.1; -  
 DR Wormpep; F40B5.2; CEO1936.  
 DR InterPro; IPR004843; M-protease.  
 DR InterPro; IPR004844; S/T phosphatase.  
 DR Pfam; PF00149; Metallophos; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 15 35 POTENTIAL.  
 FT TRANSMEM 129 149 POTENTIAL.  
 FT TRANSMEM 156 176 POTENTIAL.  
 FT TRANSMEM 205 225 POTENTIAL.  
 FT TRANSMEM 416 436 POTENTIAL.  
 SQ SEQUENCE 614 AA; 68990 MW; 20BBB8AAB524EDDB CRC64;

Query Match 48.1%; Score 38; DB 1; Length 614;  
 Best Local Similarity 42.9%; Pred. No. 21;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 NWIKLKMPTVYAGY 14  
 Db 15 NWIKLKMPTVYAGY 28

## RESULT 9

VP40 HSVEB STANDARD; PRT; 646 AA.  
 AC P28936; 069263; Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Capsid protein VP40 (Virion structural gene 35 protein) [Contains:  
 DE Capsid protein VP24 (Assemblin) (Protease) (EC 3.4.21.97); Capsid  
 DE protein VP22a].  
 GN 35.  
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 OX NCBI\_TaxID=31520;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92295566; PubMed=3138606;  
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
 RT "The DNA sequence of equine herpesvirus-1.";  
 RL Virology 189:304-316(1992).  
 CC -1- FUNCTION: VP22A IS A COMPONENT OF THE CAPSID CORE INVOLVED IN  
 CC PROCESSING AND PACKAGING OF PROGENY DNA. VP24 IS A PROTEASE  
 CC WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF AND VP22A AT THE C-  
 CC TERMINUS.  
 CC -1- CATALYTIC ACTIVITY: Cleaves -Ala-Ser- and -Ala-Ala-bonds in  
 CC the scaffold protein.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21.  
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: M86664; AA02470.1; -  
 DR EMBL: M86664; AA02471.1; -  
 DR PIR: I36798; WZBEC8.  
 DR HSSP: P16753; WZBEC8.  
 DR MEROPS; S21.001; -  
 DR InterPro; IPR001847; Assemblin.  
 DR Pfam; PF00716; Peptidase S21; 1.  
 DR PRINTS; PR00236; HSVCAPSIDP40.  
 KW Coat protein; Hydrolase; Serine protease.  
 FT CHAIN 1 646 GENE 35 PROTEIN.  
 FT CHAIN 318 646 GENE 35.5 PROTEIN.  
 FT CHAIN 1 242 COAT PROTEIN VP24 (PROTEASE).  
 FT CHAIN 243 7622 COAT PROTEIN VP22A.

FT PROPEP 7623 646 C-TERMINAL PEPTIDE.  
 FT SITE 242 243 CLEAVAGE (BY THE PROTEASE) (PROBABLE).  
 FT SITE 622 623 CLEAVAGE (BY THE PROTEASE) (PROBABLE).  
 FT ACT\_SITE 55 55 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 646 AA; 68579 MW; 1C1C28C848116276 CRC64;

Query Match 48.1%; Score 38; DB 1; Length 646;  
 Best Local Similarity 50.0%; Pred. No. 22;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LKMPVYAGY 14  
 Db 11 VSLPIVAGY 20

## RESULT 10

K2C5\_BOVIN STANDARD; PRT; 166 AA.  
 AC P04262;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Keratin, type II cytoskeletal 68 kDa, component IB (fragment).  
 OS Bos taurus (bovine).  
 CC Buxatryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85128114; PubMed=6084625;  
 RA Jorcano J.L., Franz J.K., Franke W.W.;  
 RT "Amino acid sequence diversity between bovine epidermal cyokeratin  
 RT polypeptides of the basic (type II) subfamily as determined from cDNA  
 RT clones.";  
 RL Differentiation 28:155-163(1984).  
 CC -1- SUBUNIT: HETEROTRIMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
 CC MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)  
 CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: K03534; AAA30601.1; -  
 DR PIR: A02948; KRB02B.  
 DR InterPro; IPR001664; IF.  
 DR Pfam; PF00038; filament; 1.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Keratin.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 41 .ROD.  
 FT DOMAIN 42 166 .TAIL.  
 FT DOMAIN 41 41 COIL 2B.  
 SQ SEQUENCE 166 AA; 15505 MW; 4BPE45A7C6B4BEO CRC64;

Query Match 46.8%; Score 37; DB 1; Length 166;  
 Best Local Similarity 46.7%; Pred. No. 8.3;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NWIKLKMPTVYAGY 15  
 Db 17 NWIKLKMPTVYAGY 31

```

RESULT 11
K2C4 BOVIN STANDARD; PRT; 182 AA.
AC P04260;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II cytoskeletal 55 kDa, component IV (Keratin K6 gamma)
DE (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85128114; PubMed=6084625;
RA Jorcano J.L., Franz J.K., Franke W.W.;
RT "Amino acid sequence diversity between bovine epidermal cytoke-
RT polypeptides of the basic (type II) subfamily as determined from cDNA
RT clones."
RL Differentiation 28:155-163(1984).
RN [2]
RP SEQUENCE OF 99-182 FROM N.A.
RA Navarro M., Segreles C., Jorcano J.L.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: HETEROETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFILAMENTAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03536; AAA30603.1; -
DR EMBL; Z48499; CA848399.1; -
DR PIR; A02946; A02946.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT NON TER 1 1
FT DOMAIN 1 83 ROD.
FT DOMAIN 84 182 TAIL.
FT DOMAIN 84 182 COIL 2.
FT SITE 27 27 STUTTER.
SQ SEQUENCE 182 AA; 18670 MW; 740AF9906AE3B66C CRC64;

Query Match 46.8%; Score 37; DB 1; Length 182;
Best Local Similarity 46.7%; Pred. No. 9.1;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNVKLADVEIATYR 77
DB 63 MNVKLADVEIATYR 77

RESULT 12
K2CA BOVIN STANDARD; PRT; 182 AA.
AC P04263;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II cytoskeletal 68 kDa, component IA (Fragment).
DE Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85128114; PubMed=6084625;
RA Jorcano J.L., Franz J.K., Franke W.W.;
RT "Amino acid sequence diversity between bovine epidermal cytoke-
RT polypeptides of the basic (type II) subfamily as determined from cDNA
RT clones."
RL Differentiation 28:155-163(1984).
CC -1- SUBUNIT: HETEROETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFILAMENTAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03533; AAA30599.1; -
DR PIR; A02949; KRB02A.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT NON TER 1 1
FT DOMAIN 1 66 ROD.
FT DOMAIN 67 182 TAIL.
FT DOMAIN 67 182 COIL 2B.
FT DOMAIN 87 86 H2 SUBDOMAIN.
FT DOMAIN 87 162 E2 SUBDOMAIN.
FT DOMAIN 163 182 STUTTER.
FT SITE 4 4
SQ SEQUENCE 182 AA; 18135 MW; 43C33CAE17F23287 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 182;
Best Local Similarity 46.7%; Pred. No. 9.1;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNVKLADVEIATYR 15
DB 42 MNVKLADVEIATYR 56

RESULT 13
Y063 HUMAN STANDARD; PRT; 202 AA.
AC Q15040;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein KIAA0063 (HA1234).
GN KIAA0063.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (K1A0041-K1A0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229(1994).

```

[2]  
 SEQUENCE FROM N.A.  
 MEDLINE=20057165; PubMed=10591208;  
 Dunham I., Hunt A.R., Collins J.E.,  
 Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
 Bagunley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 Cleeg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 Evans K.L., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,  
 Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 Laidlaw G.K., Langford C.F., Leverisha M.A., Lloyd A.M., Lloyd D.M.,  
 Matyln I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 Mcclay J., McLaren S., Mcmurray A.A., Milne S.L., Mortimore B.J.C.T.,  
 Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,  
 Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 Soderlund C., Spragon L., Steward C.A., Suleston J.E., Swann R.M.,  
 Vandin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilting L.,  
 Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,  
 Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenyon S., Lai H., Lao H.I.,  
 Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,  
 Wang C., Wang Y., Wang Z., White J., White J., Miller N., Mink P.,  
 Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 Hands K., Kemp K., Latreille P., Layman D., Ozeresky P., Rohlfing T.,  
 Schaefer P., Walker C., Wamsley A., Woldman P., Pepin K., Nelson J.,  
 Korfi I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 Emanuel B.S., Shaikh T., Kurahashi H., Salta S., Budarf M.L.,  
 McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyraud M., Kedra D.,  
 Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,  
 Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 Tiliun Y., Wright H.,  
 "The DNA sequence of human chromosome 22."  
 Nature 402:489-495(1999).  
 [3]  
 SEQUENCE FROM N.A.  
 TISSUE=Skin;  
 Strausberg R.;  
 Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; D31884; BAA06682.1; -  
 CC EMBL; AL021707; CAB42863.1; -  
 CC EMBL; BC015026; AAH15026.1; -  
 CC InterPro; IPR002950; Josephin.  
 CC Pfam; PF02099; Josephin; 1.  
 CC Hypothetical protein; Transmembrane.  
 CC TRANSMEM 107 129 POTENTIAL.  
 CC SEQUENCE 202 AA; 23198 MW; 837268C93C0A5A9 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 202;  
 Best Local Similarity 53.8%; Pred. No. 10;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNKIKMPMYIAGK 13  
 DB 153 LPSKIKMPWIGG 165  
 RESULT 14  
 K2C8 BOVIN STANDARD; PRT; 370 AA.  
 AC P05786;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Keratin, type II cytoskeletal 8 (Cytokeratin 8) (Cytokeratin A)  
 OS Bos taurus (bovine)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NX NCBI\_TaxID=9913;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=86193259; PubMed=2422084;  
 Magin T.M., Jorcano J.L., Franke W.W.;  
 "Cytokeratin expression in simple epithelia. II. cDNA cloning and  
 sequence characteristics of bovine cytokeratin A (no. 8)."  
 Differentiation 30:254-264(1986).  
 CC -1- SUBUNIT: HETEROHEPTAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC KERATIN 8 ASSOCIATES WITH KERATIN 18.  
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
 CC MICROFILAMENTAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)  
 CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; K03532; AAA30598.1; -  
 CC EMBL; X12877; CAA31370.1; -  
 CC PIR; A25004; A25004.  
 CC InterPro; IPR001664; IF.  
 CC Pfam; PF00038; Filament; 1.  
 CC PROSITE; PS00226; IF; 1.  
 CC KM Intermediate filament; Coiled coil; Keratin.  
 CC FT NON TER 1  
 CC FT DOMAIN <1 297 ROD.  
 CC FT DOMAIN 298 370 TAIL.  
 CC FT DOMAIN <1 25  
 CC FT DOMAIN 26 42 COIL 1A.  
 CC FT DOMAIN 43 134 COIL 1B.  
 CC FT DOMAIN 135 158 LINKER 13.  
 CC FT DOMAIN 159 297 COIL 2.  
 CC FT SITE 241 241 STUTTER.  
 CC SEQUENCE 370 AA; 42395 MW; E1PFC2782675MD3 CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 370;  
 Best Local Similarity 46.7%; Pred. No. 19;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MNKIKMPMYIAGK 15  
 DB 277 MNVKALDVEIATYR 291  
 RESULT 15  
 K2CD HUMAN STANDARD; PRT; 384 AA.  
 ID K2CD HUMAN  
 AC P48677;  
 DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Keratin, type II cytoskeletal 6D (Cytokeratin 6D) (CK 6D) (K6D)  
 DE Keratin (Fragment).  
 GN KRT6D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=95355491; PubMed=7543104;  
 RA Takahashi K., Paladini R.D., Coulombe P.A.;  
 RT "Cloning and characterization of multiple human genes and cDNAs  
 encoding highly related type II keratin 6 isoforms."  
 RL J. Biol. Chem. 270:18581-18592(1995).  
 CC -1- FUNCTION: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFILLLAR  
 CC KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II (NEUTRAL TO  
 CC BASIC; 56-70 kDa) [K1 TO K8]. BOTH A BASIC AND AN ACIDIC KERATIN  
 CC ARE REQUIRED FOR FILAMENT ASSEMBLY.  
 CC -1- SUBUNIT: HETERODIMER OF A TYPE I AND A TYPE II KERATIN. KRT6  
 CC ISOMERS ASSOCIATE WITH KRT16 AND/OR KRT17.  
 CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN DISTINCT TYPES  
 CC OF EPITHELIA SUCH AS THOSE IN ORAL MUCOSA, ESOPHAGUS, PAPILLAE  
 CC OF TONGUE AND HAIR FOLLICLE OUTER ROOT SHEATH.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST SIX ISOFORMS OF HUMAN TYPE II  
 CC KERATIN 6 (K6), K6A BEING THE MOST ABUNDANT REPRESENTING ABOUT 77%  
 CC OF ALL FORMS FOUND IN EPITHELIA.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L42610; AAB60696.1; -  
 DR EMBL; L42603; AAB60696.1; JOINED.  
 DR EMBL; L42604; AAB60696.1; JOINED.  
 DR EMBL; L42605; AAB60696.1; JOINED.  
 DR EMBL; L42606; AAB60696.1; JOINED.  
 DR EMBL; L42607; AAB60696.1; JOINED.  
 DR EMBL; L42608; AAB60696.1; JOINED.  
 DR EMBL; L42609; AAB60696.1; JOINED.  
 DR InterPro: IPR001664; IF.  
 DR Pfam: PRO0038; filament; 1.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Keratin; Multigene family.  
 FT NON\_TER 1  
 FT DOMAIN <1 292 ROD.  
 FT DOMAIN 293 384 TAIL.  
 FT DOMAIN <1 18 COIL\_1A.  
 FT DOMAIN 19 37 LINKER\_1.  
 FT DOMAIN 38 129 COIL\_1B.  
 FT DOMAIN 130 153 LINKER\_12.  
 FT DOMAIN 154 292 COIL\_2.  
 FT SITE 234 234 STUTTER.  
 SQ SEQUENCE 384 AA; 42468 MW; D69DCFD808C94E91 CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 384;  
 Best Local Similarity 46.7%; Pred. No. 20;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds

(without alignments)  
243,919 Million cell updates/sec

Title: US-09-142-524d-30

Sequence: 1 MNILKMPMTAGYK 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP rivirus:\*  
16: SP bacteriap:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	374	10	Q8RUR1
2	71	89.9	367	10	Q93XL6
3	70	88.6	367	10	Q91LT2
4	70	88.6	367	10	Q91LT1
5	70	88.6	367	10	Q9M4S6
6	70	88.6	367	10	Q9M4S4
7	70	88.6	367	10	Q9M4S3
8	70	88.6	367	10	Q9M4S2
9	70	88.6	367	10	Q93XS1
10	66	83.5	367	10	Q9M4S5
11	44	55.7	1202	5	Q21303
12	43	54.4	345	5	Q21325
13	43	54.4	645	10	Q941V8
14	43	54.4	965	10	Q882E8
15	42	53.2	939	5	O62002
16	41	51.3	252	16	Q971M1

17	41	51.9	1017	5	Q9VM32	Q9VM32 drosophila
18	41	51.9	1030	5	Q9GZG7	Q9GZG7 drosophila
19	41	51.9	1086	5	Q9VM31	Q9VM31 drosophila
20	40	50.6	184	17	Q977C2	Q977C2 sulfobolus
21	40	50.6	330	17	Q8TL23	Q8TL23 methanosarc
22	40	50.6	357	5	O62005	O62005 branchiost
23	40	50.6	414	16	O8XM91	O8XM91 clostridium
24	40	50.6	522	10	Q9LFX0	Q9LFX0 arabidopsis
25	40	50.6	617	11	Q9D2K8	Q9D2K8 mus musculus
26	40	50.6	650	2	Q9S6S6	Q9S6S6 lactococcus
27	39	49.4	162	16	Q9JWU3	Q9JWU3 neisseria m
28	39	49.4	246	2	Q9X1A1	Q9X1A1 bacteroides
29	39	49.4	294	8	Q02676	Q02676 podospira a
30	39	49.4	350	5	O62001	O62001 branchiost
31	39	49.4	395	16	Q8ZA28	Q8ZA28 yeastina pe
32	39	49.4	528	2	O52556	O52556 pseudomonas
33	39	49.4	757	5	Q21407	Q21407 caenorhabdi
34	38	48.1	227	8	Q8WD75	Q8WD75 lepidosiren
35	38	48.1	268	16	Q9RC96	Q9RC96 bacillus ha
36	38	48.1	320	16	Q9CPA7	Q9CPA7 lactococcus
37	38	48.1	335	5	O62000	O62000 branchiost
38	38	48.1	365	5	O61999	O61999 branchiost
39	38	48.1	378	13	Q90WPO	Q90WPO trachemys s
40	38	48.1	444	5	Q9V639	Q9V639 drosophila
41	38	48.1	455	5	O8ST16	O8ST16 caenorhabdi
42	38	48.1	482	5	O8ST17	O8ST17 caenorhabdi
43	38	48.1	541	16	Q9KUB4	Q9KUB4 vibrio chol
44	38	48.1	551	4	O95678	O95678 homo sapien
45	38	48.1	551	4	Q9NSA9	Q9NSA9 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q8RUR1 PRELIMINARY; PRT; 374 AA.  
AC Q8RUR1;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Cry j 1 precursor.  
GN CRY j 1.1 OR CRY j 1.2.  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta.  
OC Spermatophyta; Coniferales; Cupressaceae; Cryptomeria.  
OX NCBI\_TaxID=3369;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=POLLEN;  
RC Futamura N., Shinohara K.;  
RT "Isolation and characterization of cDNAs encoding major allergen Cry j  
RT 1 from Cryptomeria japonica pollen.";  
RU Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB081309; BAB86286.1; -;  
DR EMBL; AB081310; BAB86287.1; -;  
KW Signal.  
FT SIGNAL.  
FT CHAIN  
SQ SEQUENCE 374 AA; 40720 MW; 90D0085D24BF2BD4 CRC64;

Query Match 100.0%; Score 79; DB 10; Length 374;  
Best Local Similarity 100.0%; Pred. No. 1.8e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MNILKMPMTAGYK 15  
|||  
Db 97 MNILKMPMTAGYK 111

RESULT 2  
Q93XL6 PRELIMINARY; PRT; 367 AA.

AC 093XL6;\* (TREMBlrel. 19, Created).  
 DT 01-DEC-2001 (TREMBlrel. 19, Created).  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative allergen Cup A 1 precursor.  
 GN CUP A 1.  
 OS Cupressus arizonica.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=49011;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLLEN;  
 RA Buteroni C., Di Felice G., Pini C.;  
 RT "Cloning of Cupressus arizonica major allergen."  
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL, AJ278498; CAC37790.2; -  
 KW Signal.  
 FT CHAIN 22 21 POTENTIAL.  
 SQ SEQUENCE 367 AA; 39809 MW; AAF97260423A9F28 CRC64;  
 Query Match 89.9%; Score 71; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 5.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLMPMTIAGYK 15  
 DB 97 MNIKLMPMTIAGYK 111

RESULT 3  
 Q9LIT2 PRELIMINARY; PRT; 367 AA.  
 ID Q9LIT2;  
 AC Q9LIT2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen wajor allergen 1-2.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21315424; PubMed=11422137;  
 RX Midoro-Horitsu T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana)."  
 RL Clin. Exp. Allergy 31:771-778 (2001).  
 DR EMBL, AF151427; AAF80164.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39768 MW; 0A6AC2F1BAF89586 CRC64;  
 Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLMPMTIAGYK 15  
 DB 97 MNIKLMPMTIAGYK 111

RESULT 4  
 Q9LIT1 PRELIMINARY; PRT; 367 AA.  
 ID Q9LIT1;  
 AC Q9LIT1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-1.

OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21315424; PubMed=11422137;  
 RA Midoro-Horitsu T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana)."  
 RL Clin. Exp. Allergy 31:771-778 (2001).  
 DR EMBL, AF151429; AAF80166.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39708 MW; DCBD1981A74F4711 CRC64;  
 Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLMPMTIAGYK 15  
 DB 97 MNIKLMPMTIAGYK 111

RESULT 5  
 Q9M4S6 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S6;  
 AC Q9M4S6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monnaie R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF257491; AAF72625.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39835 MW; B4E9C60108C2C5A3 CRC64;  
 Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLMPMTIAGYK 15  
 DB 97 MNIKLMPMTIAGYK 111

RESULT 6  
 Q9M4S4 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S4;  
 AC Q9M4S4;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257493; AAF72627.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39934 MW; 974D3011D74E3D6E CRC64;

Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPLYAGYK 15  
 DB 97 MNIKLKMPLYAGYK 111  
 |||||:|:|:|

## RESULT 7

ID Q9M4S3 PRELIMINARY; PRT; 367 AA.

AC Q9M4S3; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSI.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=31469;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39832 MW; BSDPBF5A61C07A53 CRC64;

Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPLYAGYK 15  
 DB 97 MNIKLKMPLYAGYK 111  
 |||||:|:|:|

ID Q9M4S2 PRELIMINARY; PRT; 367 AA.

AC Q9M4S2; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSI.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=31469;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257495; AAF72629.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39819 MW; AAF7E055A61C07A53 CRC64;

Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPLYAGYK 15  
 DB 97 MNIKLKMPLYAGYK 111  
 |||||:|:|:|

## RESULT 9

ID Q93X51 PRELIMINARY; PRT; 367 AA.

AC Q93X51; 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Putative allergen jun o 1.  
 GN JUN O 1.  
 OS Juniperus oxycedrus (Prickly juniper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxId=69008;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=POLLEN;  
 RC Iacovacci P., Di Felice G., Pini C.;  
 RA "Cloning of Juniperus oxycedrus major allergen."  
 RT Submitted (AVG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ293767; CAC48400.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 SQ SEQUENCE 367 AA; 39808 MW; 5D28204DBFD1B9D7 CRC64;

Query Match 88.6%; Score 70; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 8.6e-05;  
 Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPLYAGYK 15  
 DB 97 MNIKLKMPLYAGYK 111  
 |||||:|:~|:|

## RESULT 10

ID Q9M4S5 PRELIMINARY; PRT; 367 AA.

AC Q9M4S5; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSI.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=31469;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257492; AAF72626.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39894 MW; 5D56FC0E3263B741 CRC64;

Query Match 83.5%; Score 66; DB 10; Length 367;  
 Best Local Similarity 73.3%; Pred. No. 0.00048;  
 Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNKMPMTIAGYK 15  
 |||||:|:|:|:  
 Db 97 MNKMPMTIAGYK 111

## RESULT 11

ID 021303 PRELIMINARY; PRT; 1202 AA.

AC 021303;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE K07G5.3 protein.  
 GN K07G5.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurtry A.A.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99069613; PubMed=9851916;  
 RX none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018 (1998).  
 DR EMBL: 271264; CA95830.1; -  
 DR InterPro: IPR000008; C2.  
 DR SMART: SM00239; C2; 1.  
 DR SEQUENCE 1202 AA; 138469 MW; 6531124389BDD063 CRC64;

Query Match 55.7%; Score 44; DB 5; Length 1202;  
 Best Local Similarity 53.8%; Pred. No. 20;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 IKKMPMTIAGYK 15  
 |||||:|:|:|:  
 Db 811 IRLNIPLYISNYK 823

## RESULT 12

ID 023625 PRELIMINARY; PRT; 345 AA.

AC 023625;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
 DE ZK829.8 protein.  
 GN ZK829.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Harris B.R.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99069613; PubMed=9851916;  
 RX none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018 (1998).  
 DR EMBL: 273899; CA98078.1; -

DR InterPro: IPR003002; 7TM\_Chemol.  
 DR InterPro: IPR000168; 7TM\_nematode.  
 DR Pfam: PF01461; 7TM\_4; 1.  
 DR SEQUENCE 345 AA; 39398 MW; 025C9DBF14FD5299 CRC64;

Query Match 54.4%; Score 43; DB 5; Length 345;  
 Best Local Similarity 53.8%; Pred. No. 8.9;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 IKKMPMTIAGYK 15  
 |||||:|:|:|:  
 Db 30 LKFSPTIGCYR 42

## RESULT 13

ID 0941V8 PRELIMINARY; PRT; 645 AA.

AC 0941V8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
 DE Hypothetical 72.5 KDa protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartioideae; Oryzaceae; Oryza.  
 CX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CV. NIPONBARE;  
 RC Buehl C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,  
 RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsilira T.,  
 RA Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G., Vanaken S.E.,  
 RA Uterback T.R., Feldblyum T.V., Quackenbush J., Salzberg S.L.,  
 RA White O., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNBa0082M15 genomic sequence."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC020666; AAK43495.1; -  
 DR InterPro: IPR003656; BHD\_finger.  
 DR InterPro: IPR002048; EF-hand.  
 DR PROSITE: PS00019; EF\_HAND; UNKNOWN\_1.  
 DR KW Hypothetical protein.  
 DR SEQUENCE 645 AA; 72497 MW; B95683F84F753C51 CRC64;

Query Match 54.4%; Score 43; DB 10; Length 645;  
 Best Local Similarity 63.6%; Pred. No. 17;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 4 KKKMPMTIAGY 14  
 |||||:|:|:|:  
 Db 437 KKKMPMTIAGY 447

## RESULT 14

ID 0852E8 PRELIMINARY; PRT; 965 AA.

AC 0852E8;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE P0022F10.3 protein.  
 GN P0022F10.3.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartioideae; Oryzaceae; Oryza.  
 CX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CV. NIPONBARE;  
 RC Sasaki T., Matsumoto T., Yamamoto K.,  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0022P10.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003229; BAB89488.1; -  
 SQ SEQUENCE 965 AA; 109160 MW; F80F90584944694E CRC64;

Query Match 54.4%; Score 43; DB 10; Length 965;  
 Best Local Similarity 63.6%; Pred. No. 25;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 KTKMPYIAGY 14  
 ||| |||  
 Db 696 KTKPLPLAGY 706

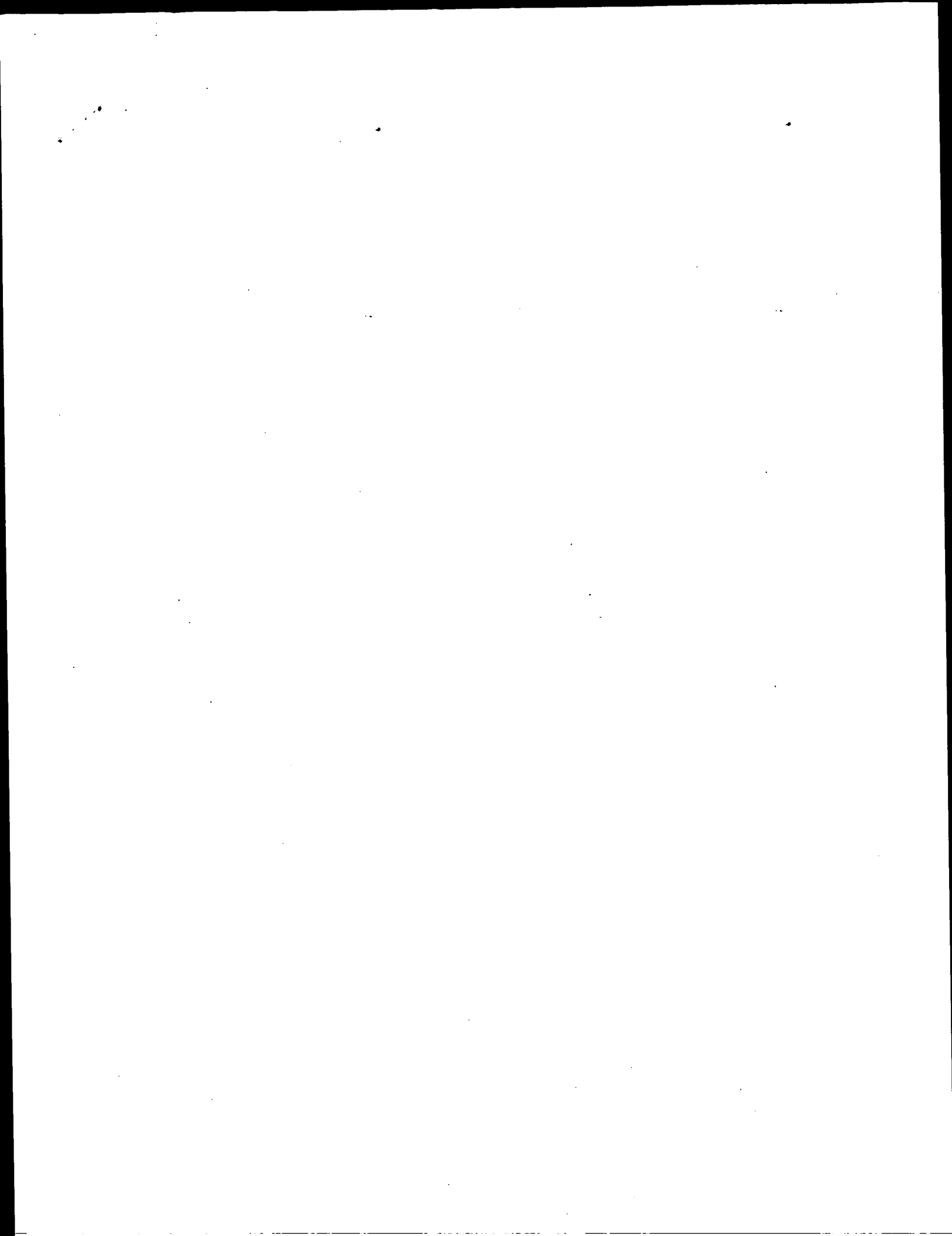
## RESULT 15

062002 PRELIMINARY; PRT; 839 AA.  
 ID 062002;  
 AC 062002;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Intermediate filament protein C1 (Fragment).  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 OX NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98267218; PubMed=9602172;  
 RA Riemer D., Karabinos A., Weber K.;  
 RT "Analysis of eight cDNAs and six genes for intermediate filament (IF)  
 RT proteins in the cephalochordate Branchiostoma reveals differences in  
 RT the IF multigene families of lower chordates and the vertebrates."  
 RL Gene 211:361-373 (1998).  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 DR EMBL: AJ223577; CAAL1444.1; -  
 DR InterPro: IPR01664; IF.  
 DR Pfam: PF00038; filament; 6.  
 DR PROSITE: PS00226; IF; 1.  
 KM Coiled coil; Intermediate filament.  
 FT NON TER 1  
 SQ SEQUENCE 839 AA; 96156 MW; 902F6DB805CAFP4 CRC64;

Query Match 53.2%; Score 42; DB 5; Length 839;  
 Best Local Similarity 53.3%; Pred. No. 33;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNITKMPYIAGYK 15  
 ||| |||  
 Db 319 MNITKLPPLIAYR 333

Search completed: April 20, 2003, 13:13:03  
 Job time : 14.6711 secs



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524d-31

Perfect score: 83

Sequence: 1 KMPWYIAGYKTFDGR 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: \*

1: PIR.73:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83	100.0	374	2 JC2124	major allergen Cry
2	83	100.0	374	2 JC2123	major allergen Cry
3	48	57.8	157	2 H87601	transcription elon
4	46	55.4	477	2 S76496	hypothetical prote
5	44	53.0	357	2 E90116	mRNA capping enzym
6	44	53.0	379	2 A70384	glutamate N-acetyl
7	43	51.8	419	2 S76571	hypothetical prote
8	42	50.6	337	2 S26863	glyceroldehydro-3-P
9	42	50.6	337	2 S40610	glyceroldehydro-3-P
10	42	50.6	337	2 T30889	valine-tRNA ligase
11	41.5	50.0	720	2 T44562	catalase (EC 1.11.
12	41	49.4	321	2 S49369	mobilitization prote
13	41	49.4	542	2 T06728	pectate lyase (EC
14	41	49.4	952	2 S38653	transposase - K1eb
15	41	49.4	1071	2 T30881	dynamin heavy chain
16	40	48.2	337	1 DEJ0GC	glyceroldehydro-3-P
17	40	48.2	368	2 G86427	probable pectate 1
18	40	48.2	974	2 A90140	conserved hypotet
19	39	47.0	162	2 E91113	hydrogenase-2 oper
20	39	47.0	162	2 E85958	member of hyb oper
21	39	47.0	162	2 F65085	hydrogenase-2 oper
22	39	47.0	410	2 T39115	formamidase-like p
23	39	47.0	463	2 B36936	nitrogenase (EC 1.
24	38	45.8	187	2 G89922	conserved hypotet
25	38	45.8	298	2 C41047	exonzyme S synthet
26	38	45.8	316	2 G72077	ct007 hypotet
27	38	45.8	316	2 G86545	alcohol dehydrogen
28	38	45.8	335	2 AC3189	pectate lyase-like
29	38	45.8	341	2 T47653	

30	38	45.8	404	2 S12209	pectate lyase (EC
31	38	45.8	404	2 A69467	hypothetical prote
32	38	45.8	418	2 T39058	probable isocitrat
33	38	45.8	529	2 A83517	probable aldehyde
34	38	45.8	568	2 T03950	probable glucose-6
35	38	45.8	605	2 T43191	probable pyruvate
36	38	45.8	688	2 H96881	protein F1E22.10 l
37	37.5	45.2	531	1 SYNCYT	tyrosine-tRNA liga
38	37	44.6	125	2 E59106	hypothetical prote
39	37	44.6	158	2 E97214	uncharacterized me
40	37	44.6	171	2 C64247	conserved hypotet
41	37	44.6	176	2 C70599	hypothetical prote
42	37	44.6	226	2 H70400	rare lipoprotein A
43	37	44.6	265	2 D82871	conserved hypotet
44	37	44.6	267	2 A75217	hypothetical prote
45	37	44.6	311	2 JC5943	catechol 1,2-dioxy

## ALIGNMENTS

## RESULT 1

JC2124  
major allergen Cry j I precursor (clone PCCI-15) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2124  
R/Source, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese

A/Reference number: JC2123; MUID:94183234; PMID:8135802

A/Accession: JC2124  
A/Molecule type: mRNA

A/Residues: 1-374 <SON>  
A/Cross-References: GB:D26545; NID:9493633; PIDN:BAA05543.1; PID:9493634

A/Experimental source: pollen  
A/Note: The authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase LRT59  
C/Keywords: glycoprotein; pollen

F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen Cry j I (clone PCCI-15) #status predicted <MAT>

F:158-191,293,354/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match Best Local Similarity 100.0%; Score 83; DB 2; Length 374;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMPWYIAGYKTFDGR 15  
DB 102 KMPWYIAGYKTFDGR 116

## RESULT 2

JC2123  
major allergen Cry j I precursor (clone PCCI-2-2) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2123; PC2065  
R/Source, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese

A/Reference number: JC2123; MUID:94183234; PMID:8135802

A/Accession: JC2123  
A/Molecule type: mRNA

A/Residues: 1-374 <SON>  
A/Cross-References: GB:D26544; NID:9493633; PIDN:BAA05542.1; PID:9493632

A/Experimental source: pollen  
A/Accession: PC2065

A/Molecule type: protein  
A/Residues: 22-53;58-81;219-232;236-258;299-307;346-372 <S02>

A/Note: The authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase LRT59  
C/Keywords: glycoprotein; pollen

F:1.21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen Cys 1 (clone PCC1-2-2) #status predicted <MAT>  
F:158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 83; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 6,6e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 15  
DB 102 KMPMYIAGYKTFDGR 116

## RESULT 3

transcription elongation factor GreA CC2848 [imported] - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C/Accession: H87601

R:Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Lamb, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: H87601

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-157 <STO>

A/Cross-references: GB:AE005673; NID:g13424460; PIDN:AAK24812.1; GSPDB:GN00148

C/Genetics:

A:Gene: CC2848

C:Superfamily: transcription elongation factor greb

C/Keywords: transcription factor

Query Match 57.8%; Score 48; DB 2; Length 157;  
Best Local Similarity 61.5%; Pred. No. 0.46;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFD 13  
DB 3 KMPMYIAGYKTFD 15

## RESULT 4

hypothetical protein - Synechocystis sp. (strain PCC 6803)

C/Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C/Accession: S76496

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
sp.

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S76496

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-477 <RXN>

A/Cross-references: EMBL:U00915; GB:AB001339; NID:g1653604; PIDN:BA18625.1; PID:g165371

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: starch synthase

Query Match 55.4%; Score 46; DB 2; Length 477;  
Best Local Similarity 46.7%; Pred. No. 3.3;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 15  
DB 79 KIPPLYFGHAFDSR 93

## RESULT 5

RNA capping enzyme [imported] - Guillardia theta nucleomorph

C/Species: nucleomorph Guillardia theta

A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C/Accession: E90116

R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re

Nature 410, 1091-1096, 2001

A/Title: The highly reduced genome of an enslaved algal nucleus.

A/Reference number: A99082; MUID:11323671; PMID:11323671

A/Accession: E90116

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-357 <DOU>

A/Cross-references: GB:AJ010592; NID:g12580792; PIDN:CAC27110.1; GSPDB:GN00151

C/Genetics:

A:Map position: 2

A/Genome: nucleomorph

C/Keywords: nucleomorph

Query Match 53.0%; Score 44; DB 2; Length 357;  
Best Local Similarity 57.1%; Pred. No. 5.6;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 14  
DB 91 KMPMYIAGYKTFDGR 104

## RESULT 6

glutamate N-acetyltransferase - Aquifex aeolicus

C/Species: Aquifex aeolicus

C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C/Accession: A70384

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: A70384

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-379 <AOB>

A/Cross-references: GB:AE000716; NID:g2983478; PIDN:AA07058.1; PID:g2983479; GB:AE0006

A/Experimental source: strain VPS

C/Genetics:

A:Gene: argu

C:Superfamily: glutamate N-acetyltransferase argu

Query Match 53.0%; Score 44; DB 2; Length 379;  
Best Local Similarity 40.0%; Pred. No. 6;

Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 15  
DB 313 KLEIYGVGYLYDVK 327

## RESULT 7

hypothetical protein - Synechocystis sp. (strain PCC 6803)

C/Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C/Accession: S76571

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

sp.



A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S76571  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-419 <KAN>  
A:Cross-references: EMBL:D64002; GB:AB001339; NID:G1001612; PIDN:BA10417.1; PID:dt01106  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match  
Best Local Similarity 51.8%; Score 43; DB 2; Length 419;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 GYKTFDGR 15  
||:|||||  
Db 183 GYKTFDGR 190

## RESULT 8

S26863

glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) - Podospora anserina  
C:Species: Podospora anserina  
C>Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 03-Jun-2002

C:Accession: S26863  
R:Ridder, R.; Osiewicz, H.D.  
Curr. Genet. 21, 207-213, 1992

A:Title: Sequence analysis of the gene coding for glyceraldehyde-3-phosphate dehydrogenase  
A:Reference number: S26863; MUID:92224295; PMID:1563046

A:Accession: S26863  
A:Molecule type: DNA  
A:Residues: 1-337 <RID>

A:Cross-references: EMBL:X62824; NID:G3101; PIDN:CAA44635.1; PID:G3102  
A:Genetics: G

A:Gene: gpd  
A:Introns: 42/3

C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase  
C:Keywords: gluconeogenesis; glycolysis; homotetramer; NAD; oxidoreductase

Query Match  
Best Local Similarity 50.6%; Score 42; DB 2; Length 337;  
Matches 8; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 3 PMYIAGY-KTFDGR 15  
||:|||||  
Db 128 PMYMGVNEKTYDGR 142

RESULT 9  
S40610

glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) - ergot fungus  
C:Species: Claviceps purpurea (ergot fungus)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 03-Jun-2002

C:Accession: S40610  
R:Jungehuelsing, U.; Arntz, C.; Smit, R.; Tudzynski, P.  
Curr. Genet. 25, 101-106, 1994

A:Title: The Claviceps purpurea glyceraldehyde-3-phosphate dehydrogenase gene: cloning,  
A:Reference number: S40610; MUID:94373854; PMID:8087877

A:Accession: S40610  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <JUN>

A:Cross-references: EMBL:X73282; NID:G452310; PIDN:CAA51721.1; PID:G452311  
A:Genetics: G

A:Introns: 42/3  
C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase  
C:Keywords: gluconeogenesis; glycolysis; oxidoreductase

Query Match  
Best Local Similarity 50.6%; Score 42; DB 2; Length 337;  
Matches 8; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 3 PMYIAGY-KTFDGR 15  
||:|||||  
Db 128 PMYMGVNEKTYDGR 142

## RESULT 10

T30889

valine-tRNA ligase (EC 6.1.1.9) - Giardia lamblia  
C:Species: Giardia lamblia  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 03-Jun-2002

C:Accession: T30889  
R:Hashimoto, T.; Sanchez, J.B.; Shirakura, T.; Muller, M.; Haegawa, M.  
Proc. Natl. Acad. Sci. U.S.A. 95, 6860-6865, 1998

A:Title: Secondary structure of mitochondria in Giardia lamblia and Trichomonas vaginalis  
A:Reference number: Z20923; MUID:96884022; PMID:9618503

A:Accession: T30889  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1218 <HAS>

A:Cross-references: EMBL:AB008525; PIDN:BA28839.1  
C:Superfamily: valine-tRNA ligase  
C:Keywords: ligase; protein biosynthesis

Query Match  
Best Local Similarity 50.6%; Score 42; DB 2; Length 1218;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 PMYIAGYKTFDGR 15  
||:|||||  
Db 465 LPARFAGYKTFDGR 478

RESULT 11  
T44562

catalase (EC 1.11.1.6) HPI [imported] - Halobacterium salinarum  
N:Alternate names: hydroperoxidase  
C:Species: Halobacterium salinarum  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C:Accession: T44562  
R:Salin, M.L.; Long, S.  
submitted to the EMBL Data Library, June 1998

A:Description: Hydroperoxidase.  
A:Reference number: Z22793

A:Accession: T44562  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-720 <SAL>

A:Cross-references: EMBL:AF069761; PIDN:AAC23534.1  
C:Superfamily: catalase HPI  
C:Keywords: heme; iron; metalloprotein; oxidoreductase

F/83/Active site: His (distal axial ligand) #status predicted  
F/248/Binding site: heme iron (His) (proximal axial ligand) #status predicted  
F/300/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match  
Best Local Similarity 50.0%; Score 41.5; DB 2; Length 720;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 1 KMPYIAGY-KTFDGR 15  
||:|||||  
Db 79 RMAWHSAGTYRFDGR 94

RESULT 12  
S49369

mobilization protein - Campylobacter coli plasmid pCCT1 and pCCT2  
C:Species: Campylobacter coli  
C>Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 08-Oct-1999

C:Accession: S49369; S49372  
R:Stonnet, V.; Quesdon, J.L.  
submitted to the EMBL Data Library, October 1994

A:Description: Nucleotide sequences and molecular organization of the two small cryptic  
A:Reference number: S49369  
A:Accession: S49369  
A:Status: preliminary  
A:Molecule type: DNA

Query Match  
Best Local Similarity 50.0%; Score 41.5; DB 2; Length 720;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 1 KMPYIAGY-KTFDGR 15  
||:|||||  
Db 79 RMAWHSAGTYRFDGR 94

RESULT 12  
S49369

mobilization protein - Campylobacter coli plasmid pCCT1 and pCCT2  
C:Species: Campylobacter coli  
C>Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 08-Oct-1999

C:Accession: S49369; S49372  
R:Stonnet, V.; Quesdon, J.L.  
submitted to the EMBL Data Library, October 1994

A:Description: Nucleotide sequences and molecular organization of the two small cryptic  
A:Reference number: S49369  
A:Accession: S49369  
A:Status: preliminary  
A:Molecule type: DNA

A/Residues: 1-321 <STO>  
 A/Cross-references: EMBL:X82080; NID:G557224; PIDN:CAA57597.1; PID:G557227  
 A/Experimental source: plasmid pCCT2  
 A/Genetics: G1  
 A/Accession: S49372  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-321 <STN>  
 A/Cross-references: EMBL:X82079; NID:G557228; PIDN:CAA57594.1; PID:G557231  
 A/Experimental source: plasmid pCCT1  
 A/Genetics: G2  
 A/Genetics: <G1>  
 A/Gene: mob  
 A/Genome: plasmid pCCT2  
 C/Genetics: <G2>  
 A/Gene: mob  
 A/Genome: plasmid pCCT1  
 C/Keywords: plasmid mobilization

Query Match 49.4%; Score 41; DB 2; Length 321;  
 Best Local Similarity 58.3%; Pred. No. 17;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTF 12  
 | | | | |  
 Db 94 KQPOYIDYKTY 105

RESULT 13  
 T06728  
 pectate lyase (EC 4.2.2.2) F28P10.100 - Arabidopsis thaliana  
 N/Alternate names: protein F28P10.100  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
 C/Accession: T06728  
 R/Querier, F.; Choisme, N.; Robert, C.; Brotier, P.; Wincker, P.; Catolico, L.; Artigou  
 submitted to the Protein Sequence Database, April 1999  
 A/Reference number: Z15793  
 A/Accession: T06728  
 A/Molecule type: DNA  
 A/Residues: 1-542 <QUB>  
 A/Cross-references: EMBL:AL049655  
 A/Experimental source: cultivar: Columbia; BAC clone F28P10  
 C/Genetics:  
 A/Map position: 3  
 A/Introns: 46/2; 346/3; 413/2; 480/2  
 A/Note: F28P10.100  
 C/Superfamily: pectate lyase LAT59  
 C/Keywords: carbon-oxygen lyase

Query Match 49.4%; Score 41; DB 2; Length 542;  
 Best Local Similarity 53.3%; Pred. No. 30;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 15  
 | | | | |  
 Db 160 KOELIINSYKTLDDGR 174

RESULT 14  
 S38653  
 transposase - Klebsiella pneumoniae  
 C/Species: Klebsiella pneumoniae  
 C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 08-Oct-1999  
 C/Accession: S38653  
 R/Rinkel, M.  
 submitted to the EMBL Data Library, November 1993  
 A/Reference number: S38652  
 A/Accession: S38653  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-952 <RIN>  
 A/Cross-references: EMBL:X75779; NID:G415974; PIDN:CAA53399.1; PID:G415976

C/Superfamily: transposase Trn21

Query Match 49.4%; Score 41; DB 2; Length 952;  
 Best Local Similarity 80.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 IAGYKTFDGR 15  
 | | | | |  
 Db 110 IYGYKTFDGR 119

RESULT 15  
 T30881  
 dynein heavy chain isotype 3A - sea urchin (Tripneustes gratilla) (fragment)  
 N/Contains: dynein ATPase (EC 3.6.4.2)  
 C/Species: Tripneustes gratilla  
 C/Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 19-Apr-2002  
 C/Accession: T30881  
 R/Gibbons, B.H.; Asai, D.J.; Tang, W.J.; Hays, T.S.; Gibbons, I.R.  
 Mol. Biol. Cell 5, 57-70, 1994  
 A/Title: Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchin  
 A/Reference number: Z20814; MUID:94243035; PMID:8186465  
 A/Accession: T30881  
 A/Status: preliminary; translated from GB/EMBL/DDBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1071 <GIB>  
 A/Cross-references: EMBL:U03970; NID:G433368; PID:G433369; PIDN:AAA63584.1  
 A/Experimental source: tissue type embryo; developmental stage mesenchyme blastula  
 C/Genetics:  
 A/Gene: DYH3A  
 C/Superfamily: dynein heavy chain, ciliary  
 C/Keywords: ATP; hydrolase; microtubule binding; P-loop

Query Match 49.4%; Score 41; DB 2; Length 1071;  
 Best Local Similarity 58.3%; Pred. No. 61;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTF 12  
 | | | | |  
 Db 1021 KLASFIAGYKSF 1032

Search completed: April 20, 2003, 13:15:44  
 Job time : 8.07895 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212,987 Million cell updates/sec

Title: US-09-142-524d-31

Perfect score: 83

Sequence: 1 KMPMTIAGYKTFDGR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83	100.0	374	1	SBP_CRYJA
2	68	83.1	346	1	MPAI_CUPAR
3	68	81.9	367	1	MPAI_JUNAS
4	60	72.3	375	1	MPAI_CHAOS
5	46	55.4	477	1	GLGA_SYNY3
6	42	50.6	337	1	G3P_GLAPU
7	42	50.6	337	1	G3P_FODAN
8	41.5	50.0	720	1	CATA_HAUNI
9	40	48.2	337	1	G3P_CRYPA
10	39	47.0	162	1	HYBE_ECOLI
11	39	45.8	316	1	VNPK_ANAVA
12	38	45.8	404	1	9612_LYCES
13	38	45.8	418	1	IDHP_SCHPO
14	38	45.8	568	1	G6PB_ORYSA
15	38	45.8	669	1	SYVM_NEUCR
16	37.5	45.2	171	1	Y428_MYCGE
17	37	44.6	226	1	RLPA_AOIAE
18	37	44.6	310	1	CTIA_ACTIW
19	37	44.6	358	1	Y993_METUA
20	37	44.6	465	1	GLGA_SYNP7
21	37	44.6	621	1	VP40_HSVBC
22	37	44.6	646	1	VP40_HSVBC
23	37	44.6	677	1	Y136_MYCTU
24	37	44.6	1254	1	MDR3_GABE1
25	37	44.6	2214	1	POLG_CRA24
26	37	44.6	2710	1	TOXA_CLADI
27	37	44.6	548	1	PCCS_MYCTO
28	36.5	44.0	752	1	CAT1_MYCTO
29	36	43.4	81	1	RS18_CHLNU
30	36	43.4	81	1	RS18_CHLNU
31	36	43.4	82	1	RS18_CHLNU
32	36	43.4	82	1	RS18_CHLNU
33	36	43.4	180	1	TRIC_TRIVU

34	36	43.4	199	1	YA70_DEIRA	Q9VFE9 deinococcus
35	36	43.4	226	1	RAN_GIALA	P38543 giardia lam
36	36	43.4	275	1	CTA2_ACTIW	O33950 actinobact
37	36	43.4	337	1	G3P_COLIN	P54117 colicetric
38	36	43.4	338	1	G3P_NEUCR	P54118 neutrospora
39	36	43.4	460	1	ANFK_RHOCA	O07935 rhodobacter
40	36	43.4	461	1	NIFN_SYNP8	O07356 synchococc
41	36	43.4	469	1	NIFN_BRAPA	P26507 bradyrhizob
42	36	43.4	506	1	PCCB_BACSU	P54541 bacillus su
43	36	43.4	541	1	60IM_HAEIN	P44973 haemophilus
44	36	43.4	549	1	PCCB_MYCLE	P53802 mycobacteri
45	36	43.4	566	1	ATP1_SCHHO	P52890 schizosacch

## ALIGNMENTS

RESULT 1  
SBP\_CRYJA  
ID SBP\_CRYJA STANDARD, PRT, 374 AA.  
AC P16632:

DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Sugi basic protein precursor (SBP) (Major allergen Cry j 1) (Cry j I).  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
OX NCBI\_TaxID=3369;

[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RP TISSUE=Pollen;  
RC MEDLINE=94183234; PubMed=8135802;  
RA Sone T., Komiyama N., Shimizu K., Kusakabe T., Morikubo K.,  
RA Kino K.,  
RT "Cloning and sequencing of cDNA coding for Cry j I, a major allergen  
RT of Japanese cedar pollen."  
RL Biochem. Biophys. Res. Commun. 199;619-625(1994).  
RN [2]

RP SEQUENCE FROM N.A.  
RC TISSUE=Pollen;  
RA Nanda M., Kuruse M., Toriogo K., Fukuda S., Kurimoto M.,  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE OF 22-41.  
RC TISSUE=Pollen;  
RX MEDLINE=89031257; PubMed=3181436;  
RA Tanai M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inoue S.,  
RA Matunasi T.,  
RT "N-terminal amino acid sequence of a major allergen of Japanese cedar  
RT pollen (Cry j I)."  
RL FEBS Lett. 239;329-332(1988).  
RN [4]

RP CARBOHYDRATES.  
RC TISSUE=Pollen;  
RX MEDLINE=95003748; PubMed=7920021;  
RA Hijioka A., Matsumoto I., Kojima K., Ogawa H.,  
RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
RT (Cryptomeria japonica) pollen allergen, Cry j I."  
RL Int. Arch. Allergy Immunol. 105;196-202(1994).  
RN [5]

RP STRUCTURE OF CARBOHYDRATES.

RC TISSUE=Pollen;  
RX MEDLINE=95332249; PubMed=7608114;  
RA Hino K., Yamamoto S., Sano O., Taniguchi Y., Kohno K., Usui M.,  
RA Fukuda S., Hanzawa H., Haruyama H., Kurimoto M.,  
RT "Carbohydrate structures of the glycoprotein allergen Cry j I from  
RT Japanese cedar (Cryptomeria japonica) pollen."  
RL J. Biochem. 117;289-295(1995).

CC -I- PTM: CONTAINS FUCOSE/XULOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
CC -I- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR  
CC POLLEN, THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
CC -I- MISCELLANEOUS: THE SEQUENCE OF CRY J I FORM A IS SHOWN HERE. FORM

B DIFFERS IN SIX POSITIONS.  
 CC ENTITIES BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D26544; BAA05542.1; -  
 CC EMBL: D26544; BAA05543.1; -  
 CC EMBL: D34639; BAA07020.1; -  
 CC PIR: A44773; A44773.  
 CC DR Glycosylated; P18632; -  
 CC DR InterPro: IPR002022; Amb\_allergen.  
 CC DR Pfam: PF00544; pec\_lyase; 1.  
 CC DR PRINTS: PRO0807; AMBALLERGEN.  
 CC Allergen; Glycoprotein; Multigene family; Signal.  
 KW  
 FT SIGNAL 1 21  
 FT CHAIN 22 374  
 FT CARBOHYD 158 158  
 FT CARBOHYD 191 191  
 FT CARBOHYD 293 293  
 FT CARBOHYD 354 354  
 FT CARBOHYD 361 361  
 FT VARIANT 12 12  
 FT VARIANT 143 143  
 FT VARIANT 202 202  
 FT VARIANT 221 221  
 FT VARIANT 358 358  
 FT VARIANT 361 361  
 FT SEQUENCE 374 AA; 40645 MW; 74AB25950248F56F CRC64;  
 SQ  
 Query Match 100.0%; Score 83; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KMPMYIAGYKTFDGR 15  
 DB 102 KMPMYIAGYKTFDGR 116  
 RESULT 2  
 MPAL\_CUPAR STANDARD; PRT; 346 AA.  
 ID MPAL\_CUPAR  
 AC Q9SGC9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major pollen allergen Cup a 1.  
 OS Cupressus arizonica.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=49011;  
 RX MEDLINE=20571526; PubMed=11122214;  
 RA Aceituno E., Del Pozo V., Minguez A., Arieta I., Cortegano I.,  
 RA Cardaba B., Gallardo S., Rojo M., Palomino P., Lahoz C.;  
 RT "Molecular cloning of major allergen from Cupressus arizonica pollen:  
 RT Cup a 1.";  
 RT Clin. Exp. Allergy 30:1750-1758(2000).  
 RL  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ243570; CAB62551.1; -  
 CC InterPro: IPR002022; Amb\_allergen.  
 CC Pfam: PF00544; pec\_lyase; 1.  
 CC DR PRINTS: PRO0807; AMBALLERGEN.  
 CC Allergen; Glycoprotein.  
 KW  
 FT CARBOHYD 127 127  
 FT CARBOHYD 157 157  
 FT CARBOHYD 272 272  
 FT SEQUENCE 346 AA; 37589 MW; F1281DCDA1D5DFD0 CRC64;  
 SQ  
 Query Match 83.1%; Score 69; DB 1; Length 346;  
 Best Local Similarity 73.3%; Pred. No. 0.00011;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 KMPMYIAGYKTFDGR 15  
 DB 81 KMPMYIAGYKTFDGR 95  
 RESULT 3  
 MPAL\_JUNAS STANDARD; PRT; 367 AA.  
 ID MPAL\_JUNAS  
 AC P81234; Q9ZNU7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major pollen allergen Jun a 1 precursor.  
 OS Juniperus ashei (Ozark white cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=1101;  
 RX MEDLINE=99414163; PubMed=10482836;  
 RA Midoro-Horiuchi T.M., Goldblum R.M., Kurosky A., Wood T.G.,  
 RA Brooks E.G.;  
 RT "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 RT allergen, Jun a 1.";  
 RT J. Allergy Clin. Immunol. 104:613-617(1999).  
 RL  
 CC -1- DISBASE: THIS PROTEIN CAUSES SEVERAL SEASONAL ALLERGENIC RHINITIS  
 CC IN NORTH AMERICA.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF106662; AAD03609.1; -  
 CC EMBL: AF106662; AAD03608.1; -  
 CC InterPro: IPR002022; Amb\_allergen.  
 CC Pfam: PF00544; pec\_lyase; 1.  
 CC DR PRINTS: PRO0807; AMBALLERGEN.  
 CC Allergen; Glycoprotein; Signal.

FT SIGNAL 1 21  
 FT CHAIN 22 367 MAJOR POLLEN ALLERGEN JUN A 1.  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 367 AA; FC9B81E75662E49 CRC64;

Query Match  
 Best Local Similarity 81.3%; Score 68; DB 1; Length 367;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 DB 102 KMPYIAGYKTFDGR 116

## RESULT 4

MEAL CHAQB STANDARD; PRT; 375 AA.  
 ID MPAL CHAQB  
 AC 096385;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major pollen allergen Cha o 1 precursor.  
 OS Chamaecyparis obtusa (Japanese cypress).  
 OC Eukaryota; Viridiplantae; Streptophyta;  
 OC Eukaryota; Viridiplantae; Coniferopsida; Coniferales; Cupressaceae;  
 OC Chamaecyparis;  
 OC NCBI\_Taxid=13415;  
 RN NCBI\_Taxid=13415;  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Pollen;  
 RX MEDLINE=96265194; PubMed=8676896;  
 RA Suzuki M., Komiyama N., Itoh H., Sone T., Kuno K., Takagi I.,  
 RA Ohta N.;  
 RT "Purification, characterization and molecular cloning of Cha o 1, a  
 RT major allergen of Chamaecyparis obtusa (Japanese cypress) pollen."  
 RT Mol. Immunol. 33:451-460 (1996).  
 RL -1. SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D45404; BAA08246.1; -  
 DR InterPro: IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR Allergen; Glycoprotein; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 375 MAJOR POLLEN ALLERGEN CHA O 1.  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 375 AA; 40258 MW; 81CD91DF7066DBF CRC64;

Query Match  
 Best Local Similarity 72.3%; Score 60; DB 1; Length 375;  
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPMYIAGYKTFDGR 15  
 DB 103 MPMYIAGYKTFDGR 116

RESULT 5  
 GLGA\_SYNY3

ID GLGA\_SYNY3 STANDARD; PRT; 477 AA.

AC P74521;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glycogen synthase (EC 2.4.1.21) (Starch [bacterial glycogen]  
 DE synthase).  
 GN GLGA OR SL0945.  
 OS Synchocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.  
 OX NCBI\_Taxid=1148;  
 RN NCBI\_Taxid=1148;  
 RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki C., Nakazaki N., Nario K.,  
 RA Okumura S., Shimo S., Takeuchi C., Wada T., Matanabe A.,  
 RA Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synchocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions."  
 RL DNA Res. 3:109-136 (1996).  
 CC -1. FUNCTION: Synthesizes alpha-1,4-glucan chains using ADP-glucose.  
 CC -1. CATALYTIC ACTIVITY: ADP-glucose + ((1,4)-alpha-D-glucosyl) (N) =  
 CC ADP + ((1,4)-alpha-D-glucosyl) (N+1).  
 CC -1. PATHWAY: Glycogen biosynthesis; second step.  
 CC -1. SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE  
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D90915; BAA18625.1; -  
 DR InterPro: IPR001296; Glycosyltransferase.  
 DR Pfam; PF00534; Glycosyltransferase; 1.  
 KW Glycogen biosynthesis; Transferase; Glycosyltransferase;  
 FT Complete proteome.  
 FT BINDING 15 15 ADP-GLUCOSE (BY SIMILARITY).  
 SQ SEQUENCE 477 AA; 54346 MW; 40C1B6013938E032 CRC64;

Query Match  
 Best Local Similarity 55.4%; Score 46; DB 1; Length 477;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 DB 79 KIPYIAGYKTFDGR 93

RESULT 6  
 G3P CLAPU STANDARD; PRT; 337 AA.

AC 000584;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).  
 GN GPD-1.  
 OS Claviceps purpurea (Ergot fungus).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocerales; Clavicipitaceae; Claviceps.  
 OX NCBI\_Taxid=5111;  
 RN NCBI\_Taxid=5111;  
 RP SEQUENCE FROM N.A.

RC STRAIN=75;  
 RX MEDLINE=94373854; PubMed=8087877;  
 RA Junghelesing U., Arntz C., Smit R., Tudzynski P.;

```

RT      "The Claviceps purpurea glyceralddehyde-3-phosphate dehydrogenase
RT      gene: cloning, characterization, and use for the improvement of a
RT      dominant selection system.";
RL      Curr. Genet. 25:101-106(1994).
CC      -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC      NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC      -1- PATHWAY: Second phase of glycolysis; first step.
CC      -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC      DEHYDROGENASE FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: X73282; CAAS1721.1; -.
DR      HSRP; P56649; IDSS.          GAP_dhydrogenase.
DR      InterPro; IPR00173;         GAP_dhydrogenase.
DR      Pfam; PF00044; gpdh. 1.
DR      Pfam; PF02800; gpdh.C; 1.
DR      PRINTS; PR00078; G3PDHCRGNASE.
DR      PROSITE; PS00071; GAPDH; 1.
KW      Glycolysis; Oxidoreductase; NAD.
FT      BINDING                     151
FT      ACT_SITE                    178      178    ACTIVATES THIOL GROUP DURING CATALYSIS.
SQ      SEQUENCE                   337 AA; 36029 MW; IC2360B9701E26B9 CRC64;

Query Match                               50.6%; Score 42; DB 1; Length 337;
Best Local Similarity                    53.3%; Pred. No. 5.6;
Matches      8; Conservative              3; Mismatches   2; Indels    2; Gaps     1;

QY      3 PMYIAGY--KTFDGR 15
DB      128 PMYVGVNKEKYDGK 142
|||||:|||||:
RESULT 7
GAP_PODAN STANDARD; PRT; 337 AA.
ID      GAP_PODAN
AC      P32637;
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).
GN      GPD.
OS      Podospira anserina.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC      Sordariales; Lasiosphaeriaceae; Podospira.
OX      NCBI_TaxID=5145;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9224295; PubMed=1563046;
RA      Ridder R., Ostlewa H.D.;
RT      "Sequence analysis of the gene coding for glyceraldehyde-3-phosphate
RT      dehydrogenase (gpd) of Podospira anserina: use of homologous
RT      regulatory sequences to improve transformation efficiency.";
RL      Curr. Genet. 21:207-213(1992).
CC      -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC      NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC      -1- PATHWAY: Second phase of glycolysis; first step.
CC      -1- SUBUNIT: HOMOTETRAMER.
CC      -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC      DEHYDROGENASE FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its

```

```
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
-----
DR      EMBL; X62824; CAA44635.1; -.
DR      PIR; S26863; S26863.
DR      HSSP; P56649; IDSS.
DR      InterPro; IPR000173; GNP_dhdrogenase.
DR      Pfam; PF00044; gpdh_1.
DR      Pfam; PF02800; gpdh_C; 1.
DR      PRINTS; PR00078; GAPDHGNASE.
DR      PROSITE; PS00071; GAPDH; 1.
DR      KMW Glycylsiste; Oxidoreductase; NAD.
FT      BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE.
FT      ACT SITE 178 178 ACTIVATES THIO GROUP DURING CATALYSIS.
SQ      SEQUENCE 337 AA; 36211 MW; C9A0FDDCC5A643B CRC64;

Query Match          50.6%; Score 42; DB 1; Length 337;
Best Local Similarity 53.3%; Pred. No. 5.6;
Matches 8; Conservative 3; Mismatches 2; Indels 2; Gaps 1.
```

```

RESULT 8
CATA_HALNT1 STANDARD;      PRT;      720 AA.
AC 073955; 09HH95;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase).
DE PERA OR VNG6294G.
OS Halobacterium sp. (strain NRC-1), and
OS Halobacterium salinarum.
OC Plasmid pNRC200.
OC Archaea, Euryarchaeota, Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxId=64091, 2242;
XX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=NRC-1; PLASMID=pNRC200;
RC MEDLINE=20504483; PubMed=110165950;
RX Ng W.V., Kennedy S.P., Mahairs G.G., Bergquist B., Pan M.,
RA Shukla H.D., Laskey S.R., Baliga N.S., Thorsson V., Sbroyna J.,
RA Swartzell S., Welt D., Hall U., Dahl T.A., Welti R., Goo Y.A.,
RA Leitnauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Meddocks D.G., Jablonski P.E., Krebs W.P., Ameghine C.M., Dale H.,
RA Iembarger T.A., Peck R.F., Daneshroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA Genomic sequence of Halobacterium species NRC-1."
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=H.salinarum;
RA Salin M.L., Long S.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1 FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).
CC -1 CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1 COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer (By
CC similarity).
CC -1 BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF005159; AAC20931.1; -  
EMBL; AF069761; AAC23534.1; -  
HSSP; P00431; IRYC.  
InterPro; IPR000763; Bac\_citase/prase.  
InterPro; IPR002016; Peroxidase.  
Pfam; PF00141; peroxidase; 1.  
PRINTS; P00458; PEROXIDASE.  
TIGR; TIGR00198; cat\_per\_HPI; 1.  
PROSITE; PS00435; PEROXIDASE\_1; 1.  
PROSITE; PS00436; PEROXIDASE\_2; 1.  
Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide; Plasmid;  
Complete proteome.  
ACT SITE 79 79 BY SIMILARITY.  
ACT SITE 83 83 BY SIMILARITY.  
METAL 248 248 IRON (PROTOME IX AXIAL LIGAND).  
SEQUENCE 720 AA; 80476 MW; FB88823BCD3CB2F9 CRC64;  
Query Match 50.0%; Score 41.5; DB 1; Length 720;  
Best Local Similarity 56.2%; Pred. No. 14;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
OY 1 KMPWYIAG-YKTFDGR 15  
DB 79 KMAHSACTYRTFDGR 94

RESULT 9  
G3P\_CRYPA STANDARD; PRT; 337 AA.  
ID G3P\_CRYPA  
AC P19089;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GPD-1).  
OS Cyphonectria parasitica (Cheesnut blight fungus) (Endothia parasitica).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Diaporthales; Valsaceae; Cyphonectria.  
OX NCBI\_TaxID=5116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91016863; PubMed=2216743;  
RA Choi G.H., Nuss D.L.;  
RT "Nucleotide sequence of the glyceraldehyde-3-phosphate dehydrogenase gene from Cyphonectria parasitica.";  
RL Nucleic Acids Res. 18:5566-5566(1990).  
CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate + NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.  
CC -1- PATHWAY: Second phase of glycolysis; first step.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X53996; CAA37943.1; -  
PIR; S11447; DEJUC.  
HSSP; P56649; IDS.  
InterPro; IPR000173; GAP\_dhndrogenase.  
Pfam; PF02800; spdh\_C; 1.  
DR Pfam; PF02800; spdh\_C; 1.

PRINTS; PR00078; G3PDHRCNASE.  
PROSITE; PS00071; GAPDH; 1.  
KW Glycolysis; Oxidoreductase; NAD.  
FT BINDING 151 151  
FT ACT SITE 178 178  
SEQUENCE 337 AA; 36194 MW; 38C804A557E68E6 CRC64;  
Query Match 48.2%; Score 40; DB 1; Length 337;  
Best Local Similarity 57.1%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 2; Indels 2; Gaps 1;  
OY 3 PMVIAGY--KTFFG 14  
DB 128 PMVWGVNERTYDG 141

RESULT 10  
HYBE\_ECOLI STANDARD; PRT; 162 AA.  
ID HYBE\_ECOLI  
AC P37183;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hydrogenase-2 operon protein hybe.  
GN HYBE OR B2992.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / TGI;  
RX MEDLINE=94292472; PubMed=8021226;  
RA Meron N.K., Chacelus C.Y., Dervartanian M., Wendt J.C.,  
RA Shanmugam K.T., Peck H.D. Jr., Przybyla A.E.;  
RT "Cloning, sequencing, and mutational analysis of the hyb operon encoding Escherichia coli hydrogenase 2.";  
RL J. Bacteriol. 176:4416-4423(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1233-1238(1997).  
CC -1- SIMILARITY: BELONGS TO THE HUPV FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U09177; AAA1593.1; -  
EMBL; U28377; AAA69159.1; -  
EMBL; AE000382; AAC76028.1; -  
Ecogene; Egi1803; hybe.  
KW Complete proteome.  
FT CONFLICT 7 7 A -> G (IN REF. 2).  
SEQUENCE 162 AA; 17977 MW; CFPD10617D815570 CRC64;  
Query Match 47.0%; Score 39; DB 1; Length 162;  
Best Local Similarity 35.7%; Pred. No. 9.3;  
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
OY 2 MPMYIAGYKTFDGR 15  
DB 38 MPMYVSDFTLFEQ 51

RESULT 11	VFNR ANAVA	STANDARD;	PRT;	463 AA.
ID	VFNR ANAVA	STANDARD;	PRT;	463 AA.
AC	057302;			
DT	15-JUN-1999 (Rel. 38, Created)			
DT	15-JUN-1999 (Rel. 38, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	(Nitrogenase vanadium-iron protein beta chain (EC 1.18.6.1)			
DE	(Nitrogenase component I) (Dinitrogenase 2 beta subunit).			
GN	VFNR.			
OS	Anabaena variabilis.			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.			
OX	NCBI_TaxID=1172;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PC 7937 / ATCC 29413;			
RX	MEDLINE=94012490; PubMed=8407800;			
RT	Thiel T;			
RT	"Characterization of genes for an alternative nitrogenase in the			
RL	Cyanobacterium Anabaena variabilis.";			
J.	Bacteriol. 175:6276-6286(1993)."			
CC	-1- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE			
CC	CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE			
CC	MOLYBDENUM-IRON PROTEIN, A VANADIUM-IRON, OR AN IRON-IRON PROTEIN.			
CC	-1- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP			
CC	= 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.			
CC	-1- COFACTOR: VANADIUM AND IRON-SULFUR.			
CC	-1- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO DELTA CHAINS.			
CC	-1- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X73824; CAA52045.1; -			
DR	EMBL; L20472; AAC36816.1; -			
DR	HSSP; P11347; IMIO.			
DR	Interpro; IPR000318; Nitrognse_comp1.			
DR	Interpro; IPR000510; Oxred_nitrognse1.			
DR	Pfam; PF00148; oxidored_nitro; 1.			
DR	PROSITE; PS00699; NITROGENASE_1.1; 1.			
DR	PROSITE; PS00090; NITROGENASE_1.2; 1.			
KW	Oxidoreductase; Nitrogen fixation; Vanadium; Iron-sulfur.			
SK	SEQUENCE 463 AA; 50833 MW; 8879442BD174563 CRC64;			
Qy	1 KMPWTIAGKTFD 13			
DB	408 KIPMYRVGFPTD 420			
Query Match		47.0%;	Score 39;	DB 1; Length 463;
Best Local Similarity		53.8%;	Pred. No. 26;	
Matches 7; Conservative		2;	Mismatches	4; Indels 0; Gaps 0
RESULT 12				
Y441.CHLPN				
ID	Y441.CHLPN	STANDARD;	PRT;	316 AA.
AC	Q928A2; Q9JRW4;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein CPN0441/CP0312/CPJ0441.			
GN	CPN0441 OR CP0312 OR CPJ0441.			
OS	Bacteria; pneumoniaceae (Chlamydotheca pneumoniaceae).			
OC	Chlamydia; Chlamydiales; Chlamydiaceae; Chlamydotheca.			
OX	NCBI_TaxID=83558;			

```

RN      SEQUENCE FROM N.A.
RP      STRAIN=CWL029;
RC      MEDLINE=99206606; PubMed=10192388;
RX      Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA      Ollinger U., Girmwood J., Davis R.W., Stephens R.S.;
RT      "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RL      Nat. Genet. 21:385-389(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AR39;
RX      MEDLINE=20150255; PubMed=10684935;
RA      Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA      White O., Hickey E.K., Peterson J., Uettersack T., Barry K., Bass S.,
RA      Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA      Gwinn M., Nelson W., DeBoy R., Kolonay J., McClary G., Salzberg S.L.,
RA      Eisen J., Fraser C.M.;
RT      "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT      pneumoniae AR39.";
RL      Nucleic Acids Res. 28:1397-1406(2000).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=J138;
RX      MEDLINE=20303049; PubMed=10871362;
RA      Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA      Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT      "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT      from Japan and CWL029 from USA.";
RL      Nucleic Acids Res. 28:2311-2314(2000).
CC      -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0441/CT007/TC0275
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL; AE001627; AAD18585.1; -.
CC      DR      EMBL; AE002194; AAF73653.1; -.
CC      DR      EMBL; AF002546; BAA98649.1; -.
CC      DR      TIGR; CP0312; -.
CC      KW      Hypothetical protein; Complete proteome.
CC      FT      CONFRACT 243 243 Y -> H (IN REF. 1).
CC      SQ      SEQUENCE 316 AA; 35415 MW; C821CB53B014B41 CRC64;
CC
CC      Query Match 45.8%; Score 38; DB 1; Length 316;
CC      Best Local Similarity 42.9%; Pred. No. 26;
CC      Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0
CC
QY      1 KMPMTYAGYKTPDG 14
      |||||: ||
      58 KLPQRTIGYQYDYG 71
      |||||: ||
      9612 LYCES
      ID      9612 LYCES STANDARD; PRT; 404 AA.
      P24336;
      DT      01-MAR-1992 (Rel. 21, Created)
      DT      01-MAR-1992 (Rel. 21, Last sequence update)
      DT      15-JUN-1999 (Rel. 38, Last annotation update)
      DE      Style development-specific protein 9612 precursor.
      GN      9612.
      OS      Lycopersicon esculentum (Tomato).
      OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      OC      Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
      OC      Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
      CX      NCBI_TaxId=4081;
      RX
      RP      SEQUENCE FROM N.A.

```



RC STRAIN=CV. V936; TISSUE=distill;  
 RX MEDLINE=91117185; PubMed=2277637;  
 RA Budelier K.A., Smith A.G., Gasser C.S.;  
 RT "Regulation of a stylar transmitting tissue-specific gene in  
 RL wild-type and transgenic tomato and tobacco.";  
 CC Mol. Gen. Genet. 224:183-192(1990).  
 CC -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPMENT OF THE TRANSMITTING  
 CC TISSUE OF THE STYLE AND/OR IN THE EVENTS RELATED TO POLLINATION  
 CC SUCH AS SOME ASPECT IN THE FACILITATION OF COMPATIBLE POLLEN TUBE  
 CC GROWTH.  
 CC -1- SUBCELLULAR LOCATION: secreted.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN THE PISTIL WHERE IT IS  
 CC FOUND IN THE OUTER FIVE LAYERS OF THE STRANDS OF TRANSMITTING  
 CC TISSUE WITHIN THE UPPER TWO-THIRDS OF THE STYLE. FOUND AT MUCH  
 CC LOWER LEVELS IN THE ANthers AND VEGETATIVE STEMS.  
 CC -1- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND DURING ANTHESIS.  
 CC -1- SIMILARITY: 54% IDENTICAL TO TOMATO PROTEINS P59 (AC P15722)  
 CC AND P56 (AC P15721).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X55193; CA38979.1; -.  
 DR F1R; S12209; S12209.  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KM Signal.  
 FT SIGNAL. 1 20  
 FT CHAIN 1 404  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 404 AA; 44298 MW; B26E69B12808675 CRC64;  
 Query Match 45.8%; Score 38; DB 1; Length 404;  
 Best Local Similarity 46.7%; Pred. No. 33;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 KMPMYTAYGKTFDGR 15  
 Db 134 KOELVWMSYTKIDGR 148  
 RESULT 14  
 IDHP SCHPO STANDARD; PRT; 418 AA.  
 AC 014254;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable isocitrate dehydrogenase [NADP], mitochondrial precursor  
 DE (EC 1.1.1.42) (Oxalosuccinate decarboxylase) (IDH) (NADP+-specific  
 DE ICDH) (IDP).  
 GN SPAC6G10.08.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OK NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 Sgourou K., Beat N., Hayes J., Baker S., Basham D., Bowman S.,  
 Brookes J., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,  
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckie E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkhardt G., Aert R., Robben J., Grymoprez B.,  
 RA Welljens I., Vanscheels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel K., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Mamdot R., Purnelle B.,  
 RA Galibert A., Aves S.J., Xiang Z., Gloux S., Lelaune V., Mottier S.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Folsburg S.L.,  
 RA Cerretti L., Lowe T., McCombie W.R., Pallesen I., Potashkin J.,  
 RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- CATALYTIC ACTIVITY. Isocitrate + NADP(+) = 2-oxoglutarate + CO(2)  
 CC + NADPH.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE  
 CC DEHYDROGENASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z98603; CAB11294.1; -.  
 DR InterPro; IPR001804; Isodh.  
 DR InterPro; IPR004790; NADp\_idh\_euk.  
 DR Pfam; PF00180; Isodh; 1.  
 DR ProDom; PD003173; NADp\_idh\_euk; 1.  
 DR TIGRfam; TIGR00127; nadp\_idh\_euk; 1.  
 DR PROSITE; PS00470; IDH IDMH; 1.  
 KM Oxidoreductase; NADP, Glyoxylate bypass; Tricarboxylic acid cycle;  
 KW Transil peptide; Mitochondrion.  
 FT TRANSIT 1 2  
 FT CHAIN 1 418  
 FT ACT SITE 104 104 PROBABLE ISOCITRATE DEHYDROGENASE [NADP].  
 FT SEQUENCE 418 AA; 47293 MW; 6E114B379C8AD61E CRC64;  
 Query Match 45.8%; Score 38; DB 1; Length 418;  
 Best Local Similarity 42.1%; Pred. No. 35;  
 Matches 8; Conservative 4; Mismatches 3; Indels 4; Gaps 1;  
 QY 1 KMPMYTAYG---YKTFDGR 15  
 Db 214 KPELVYSTKTKTKYDGR 232  
 RESULT 15  
 G6PB ORYSA STANDARD; PRT; 568 AA.  
 AC P42863;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glucose-6-phosphate isomerase, cytosolic B (EC 5.3.1.9) (GPI-B  
 DE B).  
 DE (Phosphoglucose isomerase B) (PGI-B) (Phosphohexose isomerase B) (PHI-  
 DE B).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OK NCBI\_TaxID=4530;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96293857; PubMed=8722567;
RA Nozue F., Umeda M., Nagamura Y., Minobe Y., Uchimiya H.;
RT "Characterization of cDNA encoding for phosphoglucose isomerase of
   rice (Oryza sativa L.).";
   DNA Seq. 6:127-135(1996).
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
   phosphate.
CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D45218; BAA08149.1; -.
DR HSSP; Q5N1E2; IHOX.
DR InterPro; IPR001672; G6P_Isomerase.
DR Pfam; PF00342; PGI; 1.
DR PRINTS; PR00662; G6PISOMERASE.
DR PROSITE; PS00765; P_GLUCOSE_ISOMERASE_1; 1.
DR PROSITE; PS00174; P_GLUCOSE_ISOMERASE_2; 1.
KW Gluconeogenesis; Glycolysis; Multigene family.
FT ACT_SITE 392
FT ACT_SITE 517
FT ACT_SITE 517
SQ SEQUENCE 568 AA; 62419 MW; C84C32EFE265CA12 CRC64;

```

```

Query Match 45.8%; Score 38; DB 1; Length 568;
Best Local Similarity 42.9%; Pred. No. 46;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 KMPVYIAGYKTPDG 14
   ||::|||
DB 455 KVPBHLISHKTFQG 468

```

Search completed: April 20, 2003, 13:07:34  
Job time : 3.92105 secs

GenCore version 5.1.4 p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds

(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524D-31

Perfect score: 83

Sequence: 1 KMPMYIAGYKTFDGR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virilis:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	374	10 Q8RUR1	Q8RUR1 cryptomeria
2	69	83.1	367	10 Q93XL6	Q93XL6 cupressus a
3	68	81.9	367	10 Q9LTL2	Q9LTL2 juniperus v
4	68	81.9	367	10 Q9LTL1	Q9LTL1 juniperus v
5	68	81.9	367	10 Q9M4S6	Q9M4S6 cupressus s
6	68	81.9	367	10 Q9M4S3	Q9M4S3 cupressus s
7	68	81.9	367	10 Q9M4S2	Q9M4S2 cupressus s
8	68	81.9	367	10 Q9M4S1	Q9M4S1 juniperus s
9	68	81.9	367	10 Q9M4S5	Q9M4S5 cupressus s
10	68	81.9	367	10 Q9M4S4	Q9M4S4 cupressus s
11	46	57.8	157	10 Q9A4I3	Q9A4I3 canlobacter
12	46	55.4	392	10 Q9FM66	Q9FM66 arabisopsis
13	45	54.2	122	16 Q8RE02	Q8RE02 fusobacteri
14	45	54.2	3508	5 Q9VH97	Q9VH97 diosophila
15	44	53.0	357	10 Q9AVW1	Q9AVW1 guillardia
16	44	53.0	379	16 Q67100	Q67100 aquifex aeo

17	43	51.8	419	16 Q55770	Q55770 synechocyst
18	42	50.6	690	13 Q9PVW5	Q9PVW5 oryzias lat
19	42	50.6	1218	5 Q77443	Q77443 giardia lam
20	41	49.4	181	10 Q9ZUY5	Q9ZUY5 arabidopsis
21	41	49.4	203	10 Q9M4T3	Q9M4T3 arabidopsis
22	41	49.4	206	16 Q9EWV6	Q9EWV6 streptomyces
23	41	49.4	273	12 Q91C42	Q91C42 columbid ci
24	41	49.4	273	12 Q91GAI	Q91GAI campylobact
25	41	49.4	321	2 Q57116	Q57116 campylobact
26	41	49.4	349	10 Q93E77	Q93E77 arabidopsis
27	41	49.4	417	16 Q8XYK8	Q8XYK8 ralsotonia s
28	41	49.4	501	10 Q93Z04	Q93Z04 arabidopsis
29	41	49.4	542	10 Q9SV40	Q9SV40 arabidopsis
30	41	49.4	952	2 Q48490	Q48490 klebsiella
31	41	49.4	1071	5 Q27803	Q27803 tripeustres
32	40	48.2	288	17 Q8U2S7	Q8U2S7 pyrococcus
33	40	48.2	336	16 Q8RFP1	Q8RFP1 fusobacteri
34	40	48.2	353	4 Q96OH5	Q96OH5 homo sapien
35	40	48.2	353	4 Q8WY02	Q8WY02 homo sapien
36	40	48.2	353	4 Q8WY05	Q8WY05 homo sapien
37	40	48.2	368	10 Q9C8G4	Q9C8G4 homo sapien
38	40	48.2	378	4 Q9C0J1	Q9C0J1 homo sapien
39	40	48.2	477	17 Q8TW11	Q8TW11 methanosarc
40	40	48.2	974	17 Q981D6	Q981D6 sulfolobus
41	39	47.0	162	16 Q8XBM0	Q8XBM0 escherichia
42	39	47.0	289	12 Q9Y1D4	Q9Y1D4 human echov
43	39	47.0	410	3 Q9URV7	Q9URV7 schizosacch
44	39	47.0	419	3 Q9HEV1	Q9HEV1 emericella
45	39	47.0	445	16 Q92T77	Q92T77 rhizobium m

## ALIGNMENTS

RESULT 1  
ID Q8RUR1 PRELIMINARY; PRT; 374 AA.  
AC Q8RUR1;  
DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Cry j 1 precursor.  
GN Cry j 1.1 OR Cry j 1.2.  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
OX NCBI\_TaxID=3369;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=POLLEN;  
RA Futamura N., Shinohara K.;  
RT "Isolation and characterization of cDNAs encoding major allergen Cry j  
1 from Cryptomeria japonica pollen.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB081309; BAB86286.1; -  
DR EMBL; AB081310; BAB86287.1; -  
KW Signal.  
FT SIGNAL.  
FT CHAIN.  
SQ SEQUENCE 374 AA; 40720 MW; 90D0085D24BF2BD4 CRC64;  
Query Match 100.0%; Score 83; DB 10; Length 374;  
Best Local Similarity 100.0%; Pred. No. 3.2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMPMYIAGYKTFDGR 15  
Db 102 KMPMYIAGYKTFDGR 116  
RESULT 2  
Q93XL6 PRELIMINARY; PRT; 367 AA.

```

AC 093XL6;
DT 01-DRC-2001 (TREMBlrel. 19, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative allergen Cup a 1 precursor.
GN CUP A 1.
OS Cupressus arizonica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.
OK NCBI_TaxId=49011;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RA Bitterroot C., Di Felice G., Pini C.;
RT "Cloning of Cupressus arizonica major allergen.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278498; CAC37790.2; -.
KW Signal.
FT SIGNAL. 1 21 POTENTIAL.
FT CHAIN 22 367 PUTATIVE ALLERGEN CUP A 1.
SQ SEQUENCE 367 AA; 39809 MW; AEP97260423A9F28 CRC64;

Query Match 83.1%; Score 69; DB 10; Length 367;
Best Local Similarity 73.3%; Pred. No. 0.00087;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 KMPMTIAGYKTFDGR 15
Db 102 KMPLYVAGHKTIIDGR 116

RESULT 3
O9LTL2 PRELIMINARY; PRT; 367 AA.
ID O9LTL2;
AC O9LTL2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pollen major allergen 1-2.
DS Juniperus virginiana (Eastern red cedar).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.
OK NCBI_TaxId=39584;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=21315424; PubMed=11422137;
RX Midoro-Horicut T.M., Goldblum R.M., Brooks B.G.;
RT "Identification of mutations in the genes for the pollen allergens of
RT eastern red cedar (Juniperus virginiana).";
RL Clin. Exp. Allergy 31:771-778(2001).
DR EMBL; AF151427; AA80164.1; -.
DR InterPro; IPR002022; Amb allergen.
DR Pfam; PF00544; pec_lyase; 1.
DR PRINTS; PRO0807; AMBALLERGEN.
SQ SEQUENCE 367 AA; 39768 MW; 0A6AC2F1BAF89586 CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;
Best Local Similarity 73.3%; Pred. No. 0.0013;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 KMPMTIAGYKTFDGR 15
Db 102 KMPLYVAGHKTIIDGR 116

RESULT 4
O9LTL1 PRELIMINARY; PRT; 367 AA.
ID O9LTL1;
AC O9LTL1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pollen major allergen 1-1.

```

```

OS Juniperus virginiana (Eastern red cedar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.
OK NCBI_TaxId=39584;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21315424; PubMed=11422137;
RA Midoro-Horicut T.M., Goldblum R.M., Brooks B.G.;
RT "Identification of mutations in the genes for the pollen allergens of
RT eastern red cedar (Juniperus virginiana).";
RL Clin. Exp. Allergy 31:771-778(2001).
DR EMBL; AF151429; AA80166.1; -.
DR InterPro; IPR002022; Amb allergen.
DR Pfam; PF00544; pec_lyase; 1.
DR PRINTS; PRO0807; AMBALLERGEN.
SQ SEQUENCE 367 AA; 39708 MW; DCBD1981A74E4711 CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;
Best Local Similarity 73.3%; Pred. No. 0.0013;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 KMPMTIAGYKTFDGR 15
Db 102 KMPLYVAGHKTIIDGR 116

RESULT 5
O9M4S6 PRELIMINARY; PRT; 367 AA.
ID O9M4S6;
AC O9M4S6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cup s 1 pollen allergen.
DS Cupressus sempervirens.
OS Cupressus sempervirens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.
OK NCBI_TaxId=13469;
RN (1)
RP SEQUENCE FROM N.A.
RA Monsalve R.I., Villalba M., Rodriguez R.;
RT "Cloning and expression of Cup s 1, the major allergen of the pollen
RT of Cupressus sempervirens.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257491; AAF72625.1; -.
DR InterPro; IPR002022; Amb allergen.
DR Pfam; PF00544; pec_lyase; 1.
DR PRINTS; PRO0807; AMBALLERGEN.
SQ SEQUENCE 367 AA; 39835 MW; BA89C60108C2C5A3 CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;
Best Local Similarity 73.3%; Pred. No. 0.0013;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 KMPMTIAGYKTFDGR 15
Db 102 KMPLYVAGHKTIIDGR 116

RESULT 6
O9M4S4 PRELIMINARY; PRT; 367 AA.
ID O9M4S4;
AC O9M4S4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cup s 1 pollen allergen.
DS Cupressus sempervirens.
OS Cupressus sempervirens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.
OK NCBI_TaxId=13469;

```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257493; AAF72627.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39934 MW; 974D3011D74E3D6E CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;  
 Best Local Similarity 73.3%; Pred. No. 0.0013;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 DB 102 KMPLYVAGHKTIIDGR 116

RESULT 7  
 Q9M4S3 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S3;  
 AC Q9M4S3;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSI.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39832 MW; B5DFP5A61C07A53 CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;  
 Best Local Similarity 73.3%; Pred. No. 0.0013;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 DB 102 KMPLYVAGHKTIIDGR 116

RESULT 8  
 Q9M4S2 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S2;  
 AC Q9M4S2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSI.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257495; AAF72629.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39819 MW; AE7E055A61C07A53 CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;  
 Best Local Similarity 73.3%; Pred. No. 0.0013;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 DB 102 KMPLYVAGHKTIIDGR 116

RESULT 9  
 Q93X51 PRELIMINARY; PRT; 367 AA.  
 ID Q93X51;  
 AC Q93X51;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative allergen jun o 1.  
 GN JUN O 1.  
 OS Juniperus oxycedrus (Prickly juniper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxId=69008;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLEEN;  
 RA Iacovacci P., Di Felice G., Pini C.;  
 RT "Cloning of Juniperus oxycedrus major allergen."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ293767; CAC48400.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 SQ SEQUENCE 367 AA; 39808 MW; 5D28204DBFD1B9D7 CRC64;

Query Match 81.9%; Score 68; DB 10; Length 367;  
 Best Local Similarity 73.3%; Pred. No. 0.0013;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 DB 102 KMPLYVAGHKTIIDGR 116

RESULT 10  
 Q9M4S5 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S5;  
 AC Q9M4S5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSI.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257492; AAF72626.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39894 MW; 5D56FC0E3263B741 CRC64;

Query Match 77.1%; Score 64; DB 10; Length 367;  
 Best Local Similarity 66.7%; Pred. No. 0.0064; 1; Indels 0; Gaps 0;  
 Matches 10; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 :||:|||||  
 DB 102 EMPYVAGKHTIDGR 116

## RESULT 11

Q9A413 PRELIMINARY; PRT; 157 AA.

AC 09A413;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Transcription elongation factor GreA.  
 GN CC2848.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter  
 NC NCB1\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocky J., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Hatt D.H.,  
 RA Kholony J.F., Smit J., Craven M.B., Knouri H., Shetty J., Berry K.,  
 RA Uterback T., Tran K., Wolf A., Vamathavan J., Ermoleva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AB005950; AAK24812.1; -.  
 DR HSBP; P21346; IGRJ.  
 DR TIGR; CC2848; -.  
 DR InterPro; IPR001437; GreA\_Greb.  
 DR Pfam; PF01272; GreA\_Greb; 1.  
 DR Pfam; PF03449; GreA\_Greb; 1.  
 DR PRODOM; P0004918; GreA\_Greb; 1.  
 DR PROSITE; PS00829; GREAB\_1; UNKNOWN\_1.  
 KW Elongation factor; Complete proteome.  
 SQ SEQUENCE 157 AA; 17232 MW; 880752432BA648D5 CRC64;

Query Match 57.8%; Score 48; DB 16; Length 157;  
 Best Local Similarity 61.5%; Pred. No. 1.6;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFD 13  
 :||:|||||  
 DB 3 KVMYVAGYQTLD 15

## RESULT 12

Q9FM66 PRELIMINARY; PRT; 392 AA.

AC 09FM66;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Peccate lyase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCB1\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=98290546; PubMed=9628582;

RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen  
 RT physically assigned pl and TAC clones.";  
 RL DNA Res. 5:41-54(1998).  
 DR EMBL; AB009050; BAB09239.1; -.  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; Pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 392 AA; 43682 MW; 305F94A437C30ECB CRC64;

Query Match 55.4%; Score 46; DB 10; Length 392;  
 Best Local Similarity 53.3%; Pred. No. 9.5;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 :||:|||||  
 DB 129 KQDVASVYKIDGR 143

## RESULT 13

Q8RE02 PRELIMINARY; PRT; 122 AA.

AC 08RE02;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical protein FN1336.  
 GN FN1336.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacterium.  
 NC NCB1\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=21886394; PubMed=11889109;  
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Valera O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Fongstein M., Kyridis N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586.";  
 RL J. Bacteriol. 184:2005-2018(2002).  
 DR EMBL; AE010638; AAL95532.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 122 AA; 14374 MW; 9EC1D752B888EB20 CRC64;

Query Match 54.2%; Score 45; DB 16; Length 122;  
 Best Local Similarity 53.3%; Pred. No. 3.9;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KMPYIAGYKTFDGR 15  
 :||:|||||  
 DB 9 KLSKYFGYKFNGR 23

## RESULT 14

Q9VH97 PRELIMINARY; PRT; 3508 AA.

AC 09VH97;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE CG9492 protein.  
 GN CG9492.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceilinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake R.G., Champe M., Milos G.L.G.,
RA Abiri J.F., Agbayani A., An H.-J., Andrews-Panukoch C., Baldwin D.,
RA Bailey R.M., Basu A., Baxendale J., Bayraktiroglu U., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck U., Brokstein P., Broctier P.,
RA Butts K.C., Bussan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Stretkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RA Science 287:2185-2195(2000).
DR EMBL; AE003683; AAF54422.1; -
DR FLYBase; FBgn0037726; CG9492.
DR InterPro; IPR000169; SHprot acsite.
DR PROSITE; PS00639; THIOI. PROTEASE_HIS; UNKNOWN 1.
SQ SEQUENCE 3508 AA; 402200 MW; 235A34368A30D298 CRC64;

Query Match 54.2%; Score 45; DB 5; Length 3508;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMPNYIAGYKTF 12
DB 2954 KLASFIAGYKTF 2965

RESULT 15
O9AVM1 PRELIMINARY; PRT; 357 AA.
AC O9AVM1;
DT 01-UN-2001 (TEMBLrel. 17, Created)
DT 01-UN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DB mRNA capping enzyme.
GN MCE.
OS Guillardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBT_taxonomy=55529;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20087226; PubMed=10618395;
RA Zauner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,

```

```

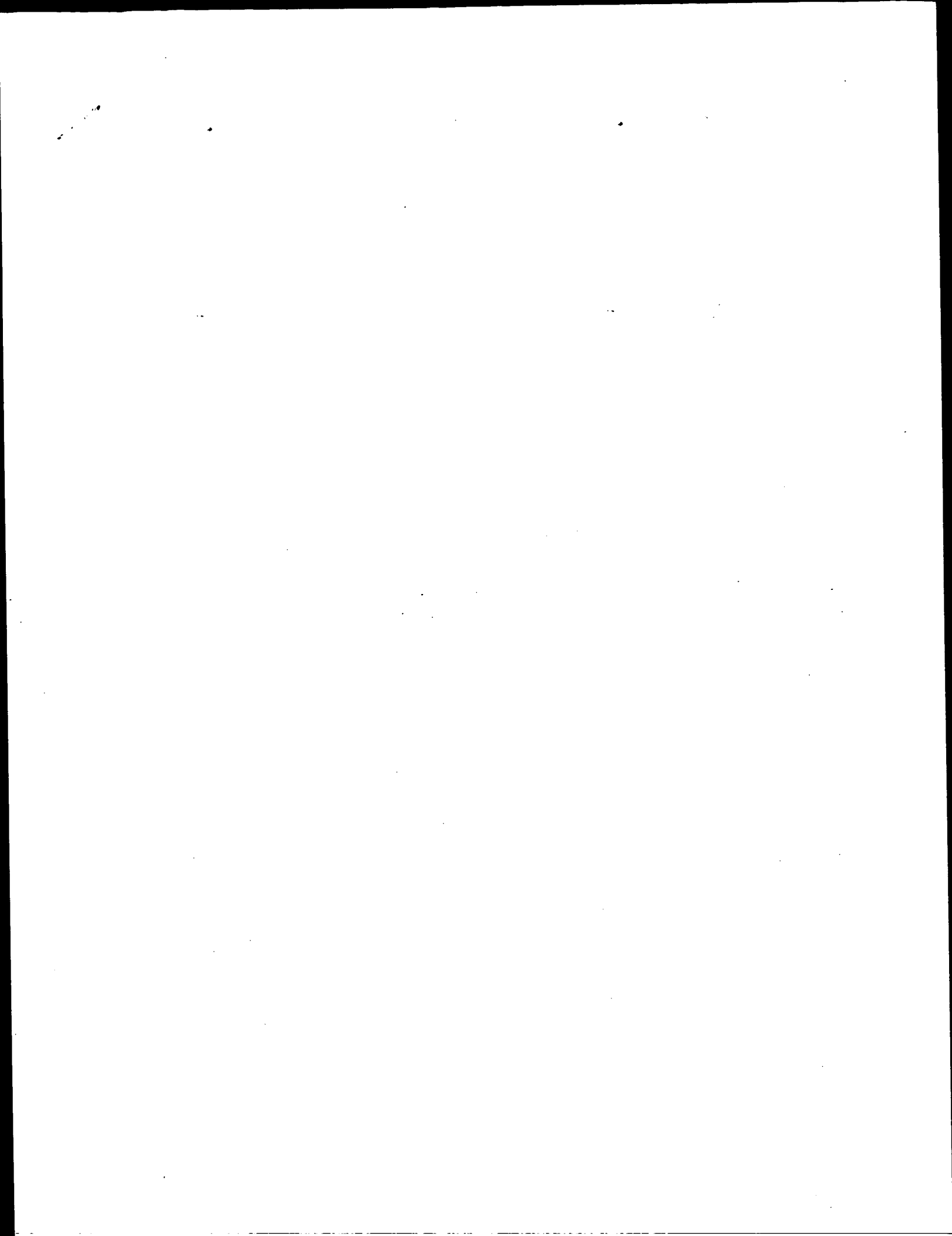
RA Cavalier-Smith T., Maier U., Douglas S.;
RT "Chloroplast protein and centromosomal genes, a tRNA intron, and odd
RT telomeres in an unusually compact eukaryotic genome, the cryptomonad
RT nucleomorph.";
RN Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RT Nature 410:1091-1096(2001).
DR EMBL; AJ010592; CAC27110.1; -
DR InterPro; IPR001339; mRNA_cap_enzyme.
DR Pfam; PF01331; mRNA_cap_enzyme; 1.
SQ SEQUENCE 357 AA; 43120 MW; 6B1BA31F717B15EB CRC64;

Query Match 53.0%; Score 44; DB 10; Length 357;
Best Local Similarity 57.1%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KMPNYIAGYKTFDG 14
DB 91 KMPNYIAGYKTFDG 104

Search completed: April 20, 2003, 13:13:05
Job time : 14.6711 secs

```





GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237,215 Million cell updates/sec

Title: US-09-142-524d-36

Perfect score: 80

Sequence: 1 PCVPIKRVSNVITIG 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR 73:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	374	2 JC2124	major allergen Cry
2	80	100.0	374	2 JC2123	major allergen Cry
3	57	71.2	390	2 H86253	hypothetical prote
4	56	70.0	394	2 T49115	pectate lyase like
5	56	70.0	394	2 T49116	pectate lyase like
6	55	68.8	404	2 T05556	pectate lyase (EC
7	55	68.8	404	2 S12209	pectate lyase (EC
8	52	65.0	368	2 G86427	probable pectate 1
9	48	60.0	374	2 T05240	pectate lyase (EC
10	48	60.0	374	2 H85148	probable pectate 1
11	48	60.0	431	2 F86179	hypothetical prote
12	46	57.5	418	2 T07701	pectate lyase (EC
13	44	55.0	463	2 T46165	pectate lyase-like
14	42	52.5	565	2 D84865	hypothetical prote
15	41	51.2	2647	2 T28161	hypothetical prote
16	40.5	50.6	528	2 T24160	hypothetical prote
17	40	50.0	256	2 D84107	leucic acid ABC
18	40	50.0	598	2 T28238	ORF MSV077 hypothe
19	40	50.0	2176	2 T39188	probable US snRNP
20	39	48.8	720	2 S50920	hypothetical prote
21	39	48.8	961	2 T23395	hypothetical prote
22	38	47.5	89	2 E53236	retrovirus-related
23	38	47.5	111	2 A72079	hypothetical prote
24	38	47.5	111	2 B86544	hypothetical prote
25	38	47.5	158	2 F65103	PTS system, n-acet
26	38	47.5	158	2 B91331	hypothetical prote
27	38	47.5	158	2 B85976	hypothetical prote
28	38	47.5	214	2 S44919	ZK688.3 protein -
29	38	47.5	394	2 G90114	FTIB related fact

30	38	47.5	587	2 A60367	transforming prote
31	38	47.5	1104	2 T01811	hypothetical prote
32	38	47.5	1328	2 S04273	retrovirus-related
33	37	46.2	71	2 T28314	ORF MSV153 hypote
34	37	46.2	79	2 B81905	hypothetical prote
35	37	46.2	254	2 E71225	hypothetical prote
36	37	46.2	274	2 T03932	pectate lyase (EC
37	37	46.2	289	2 B32459	pectate lyase (EC
38	37	46.2	289	2 S14235	Na+/K+-exchanging
39	37	46.2	290	2 A34057	Na+/K+-exchanging
40	37	46.2	290	2 A32459	Na+/K+-exchanging
41	37	46.2	290	2 J05107	Na+/K+-exchanging
42	37	46.2	315	2 C64211	thioredoxin-disulf
43	37	46.2	326	2 S51509	pectate lyase - As
44	37	46.2	327	2 T34400	hypothetical prote
45	37	46.2	421	2 H84466	hypothetical prote

## ALIGNMENTS

## RESULT 1

major allergen Cry j I precursor (clone pCCI-15) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2124

R/Source, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese

A/Reference number: JC2123; MUID:94183234; PMID:8135802

A/Accession: JC2124

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:D26545; NID:9493633; PIDN:BAA05543.1; PID:9493634

A/Experimental source: Pollen

A/Note: The authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase IAT59

C/Keywords: glycoprotein; pollen

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-374/Product: major allergen Cry j I (clone pCCI-15) #status predicted <MAT>

F:118,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 80; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. No. 1.9e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCVPIKRVSNVITIG 15  
DB 127 PCVPIKRVSNVITIG 141

## RESULT 2

major allergen Cry j I precursor (clone pCCI-2-2) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2123; PC2065

R/Source, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese

A/Reference number: JC2123; MUID:94183234; PMID:8135802

A/Accession: JC2123

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:D26544; NID:9493631; PIDN:BAA05542.1; PID:9493632

A/Experimental source: pollen

A/Accession: PC2065

A/Molecule type: protein

A/Residues: 22-53;58-81;219-232;236-258;299-307;346-372 <S02>

A/Note: The authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase IAT59

C/Keywords: glycoprotein; pollen





QY 2 CVFIRKRVSNVITIG 15  
 DB 186 CITIQVFNVIHIG 199

## RESULT 12

T07701  
 pectate lyase (EC 4.2.2.2) F17N18.100 - Arabidopsis thaliana  
 N:Alternate names: protein F17N18.100  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 23-Jul-1999  
 C:Accession: T07701  
 R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueb  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16098  
 A:Accession: T07701  
 A:Molecule type: DNA  
 A:Residues: 1-418 <BEV>  
 A:Cross-references: EMBL:AL049751  
 A:Experimental source: cultivar Columbia; BAC clone F17N18  
 C:Genetics:  
 A:Map position: 4  
 A:introns: 26/2; 49/2; 264/3; 299/2; 345/3  
 A>Note: F17N18.100  
 C:Superfamily: pectate lyase LATS9  
 C:Keywords: carbon-oxygen lyase

Query Match 57.5%; Score 46; DB 2; Length 418;  
 Best Local Similarity 50.0%; Pred. No. 2.6;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 CVFIRKRVSNVITIG 15  
 DB 175 CITVQVYVNIHIG 188

## RESULT 13

T46165  
 pectate lyase-like protein - Arabidopsis thaliana  
 N:Alternate names: protein T4D2.120  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 02-Sep-2000  
 C:Accession: T46165  
 R:Nakamura, G.; Farman, B.; Dauner, D.; Steer, W.; Holland, R.; Welcheslgartner, M.;  
 submitted to the Protein Sequence Database, December 1999  
 A:Reference number: Z23025  
 A:Accession: T46165  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-463 <NTA>  
 A:Cross-references: EMBL:AL132958  
 A:Experimental source: cultivar Columbia; BAC clone T4D2  
 C:Genetics:  
 A:Map position: 3  
 A:introns: 27/2; 112/2; 182/1; 265/3; 329/3; 399/2  
 A>Note: T4D2.120  
 C:Superfamily: pectate lyase LATS9

Query Match 55.0%; Score 44; DB 2; Length 463;  
 Best Local Similarity 53.8%; Pred. No. 6.6;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 CVFIRKRVSNVITIG 14  
 DB 159 CITIQVYVNIHIG 171

## RESULT 14

D84886  
 hypothetical protein At2g45100 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: D84886  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; McFarr, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.;  
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: D84886  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-565 <STO>  
 A:Cross-references: GB:AE002093; NID:94895242; PIDN:ADJ32827.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g45100  
 A:Map position: 2

Query Match 52.5%; Score 42; DB 2; Length 565;  
 Best Local Similarity 46.7%; Pred. No. 18;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 PCVFIRKRVSNVITIG 15  
 DB 176 PSIFIRKRVSNVITIG 190

## RESULT 15

T28161  
 hypothetical protein FCR3-varT11-1 - malaria parasite (Plasmodium falciparum) (fragments  
 C:Species: Plasmodium falciparum  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
 C:Accession: T28161  
 R:Hernandez-Rivas, R.; Mattei, D.; Sterkers, Y.; Peterson, D.S.; Wellens, T.E.; Scherf,  
 Mol. Cell. Biol. 17, 604-611, 1997  
 A:Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.  
 A:Reference number: Z20483; MUID:97154495; PMID:9001213  
 A:Accession: T28161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2647 <HBR>  
 A:Cross-references: EMBL:U67959; NID:91794255; PID:91809295; PIDN:AAC47438.1  
 A:Experimental source: strain FQ27/PNG  
 C:Genetics:  
 A:introns: 2156/3  
 A>Note: FCR3-varT11-1

Query Match 51.2%; Score 41; DB 2; Length 2647;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 CVFIRKRVSNVITIG 15  
 DB 1070 CKHVKIKIKMLIIG 1083

Search completed: April 20, 2003, 13:15:46  
 Job time : 8.07895 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524D-121

Perfect score: 79

Sequence: 1 ASKNFHLQKNTIGTG 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	15	17	AAAR97908
2	79	100.0	15	19	AAAR97908
3	79	100.0	19	22	AAAB84115
4	79	100.0	42	22	AAAB84105
5	79	100.0	80	18	AAAR27369
6	79	100.0	105	18	AAAR27370
7	79	100.0	134	18	AAAR27371
8	79	100.0	460	16	AAAR69791
9	79	100.0	514	15	AAAR53690
10	79	100.0	514	16	AAAR74333

11	79	100.0	514	16	AAAR69792	Japonicum allergen
12	79	100.0	514	17	AAAR93589	Japan cedar pollen
13	79	100.0	514	17	AAAR81586	Cedar pollen aller
14	79	100.0	514	20	AAAR25666	Japanese cedar all
15	79	100.0	514	20	AAAR25667	Japanese cedar all
16	79	92.4	17	17	AAAR1582	Japonicum allergen
17	68	86.1	17	19	AAAR80347	Sugi allergen aller
18	59	74.7	47	19	AAAR80351	Sugi allergen prot
19	59	74.7	47	19	AAAR80351	Sugi allergen prot
20	59	74.7	54	21	AAAR23904	Artificial sequenc
21	59	74.7	54	21	AAAR23904	Artificial sequenc
22	59	74.7	94	21	AAAR23905	Artificial sequenc
23	59	74.7	94	21	AAAR23905	Artificial sequenc
24	59	74.7	95	21	AAAR23907	Artificial sequenc
25	59	74.7	95	21	AAAR23907	Artificial sequenc
26	59	74.7	95	22	AAAR6908	Cedar pollen aller
27	59	74.7	96	22	AAAR69104	Cedar pollen aller
28	59	74.7	96	22	AAAR69119	Cedar pollen aller
29	59	74.7	97	22	AAAR69105	Cedar pollen aller
30	59	74.7	99	21	AAAR23902	Artificial sequenc
31	59	74.7	99	21	AAAR23907	Artificial sequenc
32	59	74.7	99	22	AAAR69099	Cedar pollen aller
33	59	74.7	214	22	AAAR69120	Cedar pollen aller
34	58	73.4	12	18	AAAR12549	Japanese cedar pol
35	58	73.4	12	22	AAAR69108	Cryptomeria japon
36	58	73.4	47	19	AAAR80352	Sugi allergen prot
37	58	73.4	47	19	AAAR80353	Sugi allergen prot
38	58	73.4	47	19	AAAR80356	Sugi allergen prot
39	58	73.4	47	19	AAAR80357	Sugi allergen prot
40	58	73.4	61	21	AAAR23883	Artificial sequenc
41	58	73.4	61	21	AAAR23889	Artificial sequenc
42	58	73.4	61	21	AAAR23894	Artificial sequenc
43	58	73.4	71	21	AAAR23878	Artificial sequenc
44	58	73.4	79	21	AAAR23880	Artificial sequenc
45	58	73.4	79	21	AAAR23884	Artificial sequenc

#### ALIGNMENTS

RESULT 1  
AAAR97908  
ID AAR97908 standard; peptide; 15 AA.  
XX  
AC AAR97908;  
XX  
DT 16-AUG-1996 (first entry)  
XX  
XX Japan cedar pollen mature allergen Cry j II amino acids 186-200.  
DE  
XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
KW Sugi pollinosis; diagnosis; treatment.  
XX  
XX Cryptomeria japonica.  
OS  
XX JP08047392-A.  
PN  
XX 20-FEB-1996.  
PD  
XX  
PF 07-NOV-1994; 94UP-0297840.  
XX  
XX 26-MAY-1994; 94UP-0134868.  
PR 05-NOV-1993; 93UP-0276773.  
XX  
XX (MEIP) MEIJI MILK PROD CO LTD.  
PA  
XX WPI; 1996-166249/17.  
DR  
XX Japan cedar pollen allergen Cry j II epitope - comprises at least  
PT part of specified 460 amino acid protein  
XX  
XX Claim 8; Fig 4; 17pp; Japanese.  
XX

CC AAR97871-R97960 are overlapping peptides used for the epitope mapping  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.

XX Sequence 15 AA;

Query Match 100.0%; Score 79; DB 17; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.3e-07; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGT 15

1 ASKNFHLQKNTIGT 15

DB 1 ASKNFHLQKNTIGT 15

RESULT 2

AAM57764

ID AAM57764 standard; peptide; 15 AA.

AC AAM57764;

DT 17-SEP-1998 (first entry)

XX Residues 186-200 of Cry j 2.

XX Cry j 2; Japanese cedar pollen antigen; allergy; immunotherapy;

XX HLA class II molecule.

XX Cryptomeria japonica.

XX WO9820902-A1.

XX 22-MAY-1998.

XX 12-NOV-1997; 97WO-JP04129.

XX 13-NOV-1996; 96JP-0302053.

XX (MEIP) MEIP MILK PROD CO LTD.

XX Daiichi K, Kuno K, Kume A, Sone T;

XX WPI; 1998-297617/26.

PT Peptides derived from Japanese cedar pollen antigens are  
 PT immunotherapeutic agents - useful for allergy treatment and typing  
 PT HLA class II molecules in allergy sufferers

PS Claim 12; Page 31; 50pp; Japanese.

XX This sequence represents residues 186-200 of the Cry j 2 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.

XX Sequence 15 AA;

Query Match 100.0%; Score 79; DB 19; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.3e-07; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGT 15

1 ASKNFHLQKNTIGT 15

DB 1 ASKNFHLQKNTIGT 15

RESULT 3

AAB84115

ID AAB84115 standard; peptide; 19 AA.

AC AAB84115;

DT 06-AUG-2001 (first entry)

XX Amino acid sequence of a pollen antigen.

XX Immunomodulatory peptide; immune response; autoimmune disease; allergy;

XX asthma; host-versus-graft rejection; T cell; anergy; apoptosis.

XX Cryptomeria japonica.

XX WO200136448-A2.

XX 25-MAY-2001.

XX 27-OCT-2000; 2000WO-US41646.

XX 27-OCT-1999; 99US-0161734.

XX (CELS-) CEL-SCI CORP.

XX Zimmerman DH;

XX WPI; 2001-374498/39.

PT Novel immunomodulatory peptide construct useful for modulating an  
 PT inappropriate immune response in an individual at risk for autoimmune  
 PT disease, allergic reactions, asthma or host-graft or graft-host disease

PS Claim 2; Page 27; 55pp; English.

XX The specification describes an immunomodulatory peptide construct.  
 CC The immunomodulatory peptide comprises a first peptide associated with  
 CC autoimmune disease, allergy or asthma, or host-versus-graft rejection  
 CC and which will bind to an antigen receptor on a set or subset of T cells,  
 CC linked to a second immune modulating peptide which will cause a directed  
 CC immune response by the set or subset of T cells to which the first  
 CC peptide is attached. Alternatively, the second peptide will bind to a  
 CC T cell receptor site on the surface of the T cell which will cause the  
 CC set or subset of T cells to initiate, but not complete, an immune  
 CC response to cause the set or subset of T cells to undergo anergy and  
 CC apoptosis. The immunomodulatory peptides are useful for eliminating a set  
 CC or subset of T cells involved in autoimmune response. They are useful for  
 CC the treatment of autoimmune disease, allergic reactions, asthma or  
 CC host-graft or graft-host rejections. The immunomodulatory peptides are  
 CC also useful for interrupting an autoimmune disease associated pathway  
 CC necessary to complete T cell activation. The present peptide is used to  
 CC construct immunomodulatory peptides of the invention, and is  
 CC representative of the first peptide.

XX Sequence 19 AA;

Query Match 100.0%; Score 79; DB 22; Length 19;

Best Local Similarity 100.0%; Pred. No. 1.7e-07; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGT 15

1 ASKNFHLQKNTIGT 15

DB 5 ASKNFHLQKNTIGT 19

RESULT 4

AAB84105



XX 12-SEP-1997.  
 PD 10-MAR-1997; 97WO-JP00740.  
 PF 10-MAR-1996; 96UP-0080702.  
 XX (MEIP) MEIJI MILK PROD CO LTD.  
 PA Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 PI WPI; 1997-470495/43.  
 DR Peptide immuno:therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 PS Claim 6; Page 31; 58pp; Japanese.  
 XX The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 XX Sequence 105 AA;

Query Match 100.0%; Score 79; DB 18; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ASKNFHLQKNTIGTG 15  
 DB 35 ASKNFHLQKNTIGTG 49

RESULT 7  
 AAM27371 ID AAM27371 standard; peptide; 134 AA.  
 XX AAM27371;  
 AC 24-MAR-1998. (first entry)  
 DT Multi-epitope peptide used as immunotherapeutic agent #3.  
 DE Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 XX T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 KW Synthetic.  
 OS WO9732600-A1.  
 PN 12-SEP-1997.  
 PD 10-MAR-1997; 97WO-JP00740.  
 PF 10-MAR-1996; 96UP-0080702.  
 XX (MEIP) MEIJI MILK PROD CO LTD.  
 PA Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 PI WPI; 1997-470495/43.  
 DR Peptide immuno:therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens

XX Claim 6; Page 32; 58pp; Japanese.  
 PS The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 XX Sequence 134 AA;

Query Match 100.0%; Score 79; DB 18; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ASKNFHLQKNTIGTG 15  
 DB 35 ASKNFHLQKNTIGTG 49

RESULT 8  
 AAR69791 ID AAR69791 standard; Protein; 460 AA.  
 XX AAR69791;  
 AC 27-SEP-1995 (first entry)  
 DT Japonicum allergen residues 55-514.  
 DE Japonicum allergen; residues 55-514; induced histamine release;  
 XX antiallergic peptide; IgE cross-linking inhibition.  
 KW Japonicum sp.  
 OS WO9502412-A.  
 PN 26-JAN-1995.  
 PD 15-JUL-1994; 94WO-JP01164.  
 PF 16-JUL-1993; 93JP-0177008.  
 PR 01-SEP-1993; 93JP-0217725.  
 PR 07-APR-1994; 94JP-0069336.  
 XX (MEIP) MEIJI MILK PROD CO LTD.  
 PA Kino K, Kohno Y, Komiyama N, Sone T;  
 PI WPI; 1995-067159/09.  
 DR N-PSDB; AAQ84044.  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 PT with IgE antibody  
 PS Disclosure; Pages 26-27; 46pp; Japanese.  
 XX AAQ84044 encodes AAR69791 Japonicum allergen residues 55-514, from  
 CC which the antiallergic peptides AAR69845-R69809 were derived.  
 CC The peptides ability to inhibit the cross-linking of an allergen,  
 CC to an IgE antibody can be used in the prevention and treatment of  
 CC allergic diseases.  
 XX Sequence 460 AA;

Query Match 100.0%; Score 79; DB 16; Length 460;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 ASKNFHLQKNTIGTG 15  
DB 186 ASKNFHLQKNTIGTG 200

## RESULT 9

AAR53690  
ID AAR53690 standard; Protein; 514 AA.  
AC AAR53690;

DT 01-FEB-1995 (first entry)

DE Japanese cedar pollen allergen Cry j II.

KW Cedar pollinosis; diagnostic.

OS Cryptomeria japonica.

PN WO9411512-A.

PD 26-MAY-1994.

PF 12-NOV-1993; 93WO-US11000.

PR 12-NOV-1992; 92US-0975179.

PA (IMMU-)\* IMMUNOLOGIC PHARM CORP.

PI Brauer A, Kuo M, Pollock J, Yeung S;

DR WPI; 1994-183513/22.

DR N-PSDB; AAQ66048.

PT Allergenic Cry j II protein and fragments from Japanese cedar pollen - used to diagnose, treat and prevent Japanese cedar pollinosis

PS Claim 2; Fig 4; 89pp; English.

CC The sequence is of a Japanese cedar pollen allergen Cry j II. The protein and its fragments can be used for diagnosis and treatment of Japanese cedar pollinosis and to identify similar sequences in other plants.  
CC See also AAR53692-6.

SQ Sequence 514 AA;

Query Match Best Local Similarity 100.0%; Score 79; DB 15; Length 514;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

## RESULT 10

AAR74333  
ID AAR74333 standard; Protein; 514 AA.

AC AAR74333;

DT 01-NOV-1995 (first entry)

DE Japanese cedar pollen allergen.

KW Japanese cedar pollen; allergen; allergy; therapy; diagnostic;

OS Cryptomeria japonica.

PN EP655500-A.

PD 31-MAY-1995.

PF 03-NOV-1994; 94EP-0308117.

PR 27-DEC-1993; 93JP-0346814.

PR 05-NOV-1993; 93JP-0299151.

PR 20-DEC-1993; 93JP-0344596.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Kurimoto M, Namba M, Torigoe K;

DR WPI; 1995-195588/26.

DR N-PSDB; AAQ90156.

PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it, useful for treatment and diagnosis of cedar pollen allergy

PS Claim 5; Page 26-28; 41pp; English.

CC The gene encoding an allergen of Japanese cedar pollen was isolated by PCR amplification using primers based on portions of the allergen protein. The gene was used for recombinant allergen production in E. coli (vector plasmid pKK-223-3).

SQ Sequence 514 AA;

Query Match Best Local Similarity 100.0%; Score 79; DB 16; Length 514;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

## RESULT 11

AAR69792  
ID AAR69792 standard; Protein; 514 AA.

AC AAR69792;

DT 27-SEP-1995 (first entry)

DE Japonicum allergen.

KW Japonicum allergen; induced histamine release; antiallergic peptide; IgE cross-linking inhibition.

OS Japonicum sp.

PN WO9502412-A.

PD 26-JAN-1995.

PF 15-JUL-1994; 94WO-JP01164.

PR 16-JUL-1993; 93JP-0177008.

PR 01-SEP-1993; 93JP-0217725.

PR 07-APR-1994; 94JP-0069336.

PA (MEIJI) MEIJI MILK PROD CO LTD.

PI Kino K, Kohno Y, Komiyama N, Sone T;

DR WPI; 1995-067159/09.

DR N-PSDB; AAQ84045, AAQ84046.

PT Peptide antiallergic agent - inhibits cross-linking of allergen with IgE antibody

PS Example 3; Pages 27-28; 46pp; Japanese.

XX AA084045 encodes AAR69792 Japonicum allergen, from which the  
CC anti-allergic peptides AAR6945-R69809 were derived. The peptides ability  
CC to inhibit the cross-linking of an allergen, to an IGE antibody can be  
CC used in the prevention and treatment of allergic diseases.

XX Sequence 514 AA;

Query Match 100.0%; Score 79; DB 16; Length 514;  
Best Local Similarity 100.0%; Pred. No. 6.7e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

RESULT 12

AAR93599 ID AAR93599 standard; Protein; 514 AA.

XX AC AAR93599;

XX DT 16-AUG-1996 (first entry)

XX DE Japan cedar pollen Cry j II allergen.

XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
XX Sugi pollinosis; diagnosis; treatment.

XX OS Cryptomeria japonica.

XX FH Key Location/Qualifiers

XX FT Peptide 1..54

XX FT /\*label= sig\_peptide

XX FT 55..514

XX FT Protein /\*label= mat\_protein

XX JP08047392-A.

XX PD 20-FEB-1996.

XX PF 07-NOV-1994; 94JP-0297840.

XX PR 26-MAY-1994; 94JP-0134868.

XX PR 05-NOV-1993; 93JP-0276773.

XX PA (MEIP) MEIUI MILK PROD CO LTD.

XX WPI; 1996-166249/17.

XX DR N-PSDB; AAT18102.

XX PT Japan cedar pollen allergen Cry j II epitope - comprises at least  
XX part of specified 460 amino acid protein

XX PS Claim 1; Page 10-11; 17pp; Japanese.

XX AAR93599 is a Japan cedar pollen Cry j II allergen which is useful  
XX in the diagnosis, prevention and treatment of Sugi pollinosis.  
XX the allergic reaction to Japan cedar pollen. Significant regions of  
XX the allergen were identified using overlapping peptides of the full  
XX epitope derived from a Cry j II antigen-specific T cell line  
XX (see AAR97871-R97960). Amino acids 66-80 (AAR97884) and 186-200  
XX (R978908) of the full mature 460 amino acid allergen are the most  
XX allergenic of the 90 peptides tested.

XX Sequence 514 AA;

Query Match 100.0%; Score 79; DB 17; Length 514;

Best Local Similarity 100.0%; Pred. No. 6.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

RESULT 13

AAR81586 ID AAR81586 standard; Protein; 514 AA.

XX AC AAR81586;

XX DT 24-MAY-1996 (first entry)

XX DE Cedar pollen allergen A.

XX Cedar; pollen; allergen; immunoglobulin E; IGE; T-cell epitope;  
XX antibody; pollinosis; therapy; immunotherapy.

XX OS Cryptomeria japonica.

XX PN EP700929-A2.

XX PD 13-MAR-1996.

XX PF 08-SEP-1995; 95EP-0306295.

XX PR 14-JUL-1995; 95JP-0200221.

XX PR 10-SEP-1994; 94JP-0242137.

XX PR 14-JUL-1995; 95JP-0200204.

XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

XX PI Hino K, Saito S, Taniguchi Y;

XX WPI; 1996-140976/15.

XX New peptide(s) derived from cedar pollen allergens - activate  
XX PT allergen-specific T-cells, but not allergen-specific IGE antibodies,  
XX used for treating cedar pollinosis

XX PS Claim 5; Page 29-30; 36pp; English.

XX Synthetic peptides based on portions of cedar pollen allergens A  
XX (AAR81586) and B (AAR81587) were tested for their ability to activate  
XX cedar allergen-specific T-cells, but not allergen-specific IGE  
XX antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
XX epitopes. These peptides, plus subsequences (AAR81573-79) essential  
XX for T-cell recognition, and homologous peptides (AAR81588-96) can  
XX be used as immunotherapeutic agents to treat or prevent cedar  
XX pollinosis, avoiding side-effects such as anaphylaxis.

XX Sequence 514 AA;

Query Match 100.0%; Score 79; DB 17; Length 514;

Best Local Similarity 100.0%; Pred. No. 6.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

RESULT 14

AAR25666 ID AAR25666 standard; Protein; 514 AA.

XX AC AAR25666;

XX DT 30-SEP-1999 (first entry)

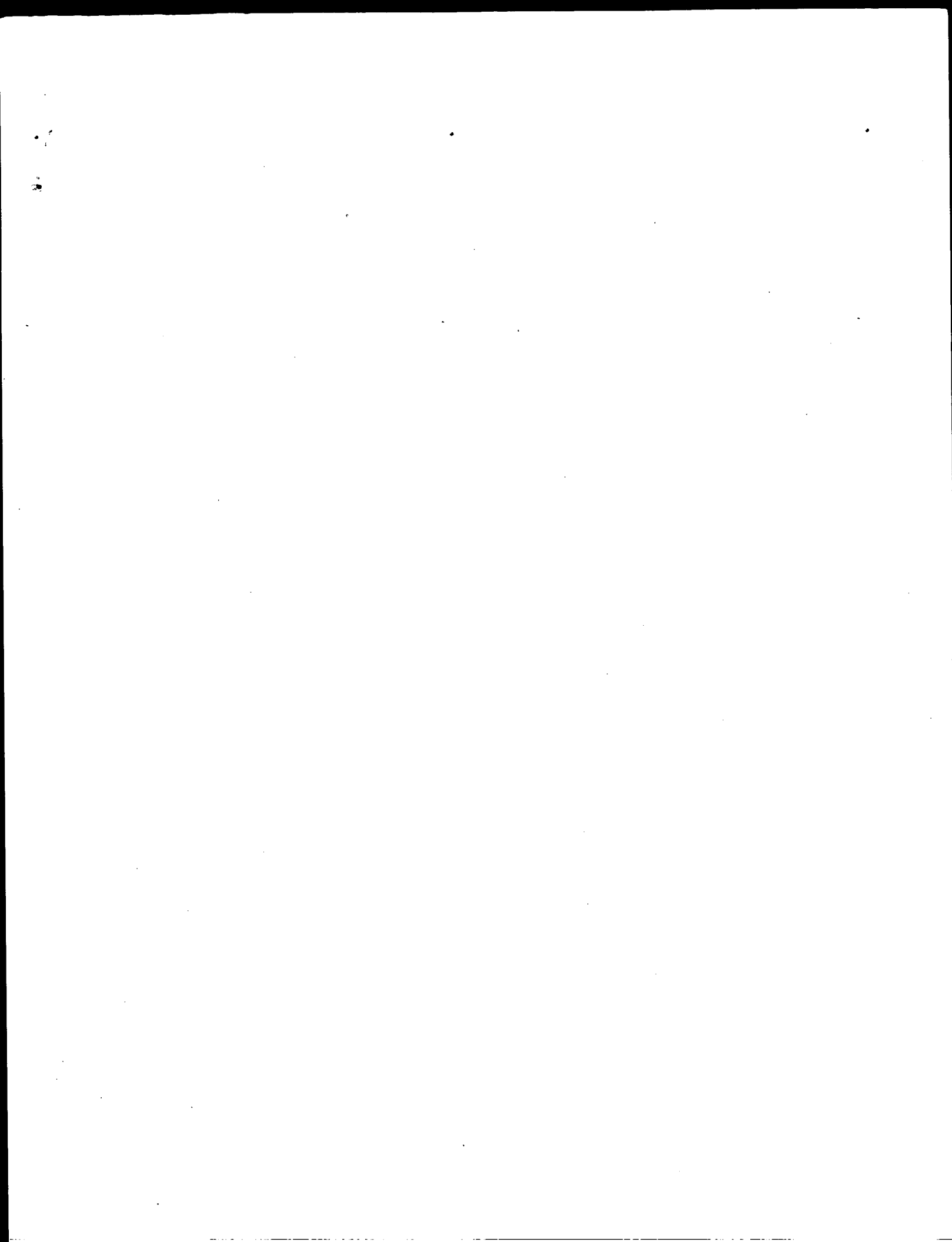
XX Japanese cedar allergen 1076242 Cry j II precursor protein fragment.

XX Major histocompatibility complex; class II; desensitising; human;

KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 OS Cedrus sp.  
 XX WO9934826-A1.  
 XX 15-JUL-1999.  
 XX 11-JAN-1999; 99WO-GB00080.  
 XX 21-SEP-1998; 98GB-0020474.  
 XX 09-JAN-1998; 98GB-0000445.  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PI Kay AB, Larche M;  
 PT WPI; 1999-458255/38.  
 PS Desensitizing patients to polypeptide allergens  
 PS Example 6; Page 74; 117pp; English.  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076242 Cry j II  
 CC precursor.  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 79; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 6,7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ASKNFHLQKNTIGTG 15  
 DB 240 ASKNFHLQKNTIGTG 254  
 RESULT 15  
 AAY25667  
 ID AAY25667 standard; protein; 514 AA.  
 AC AAY25667;  
 XX 30-SEP-1999 (first entry)  
 DT Japanese cedar allergen 1076241 Cry j II protein fragment.  
 DE Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

OS Cedrus sp.  
 XX WO9934826-A1.  
 XX 15-JUL-1999.  
 XX 11-JAN-1999; 99WO-GB00080.  
 XX 21-SEP-1998; 98GB-0020474.  
 XX 09-JAN-1998; 98GB-0000445.  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PI Kay AB, Larche M;  
 PT WPI; 1999-458255/38.  
 PS Desensitizing patients to polypeptide allergens  
 PS Example 6; Page 74; 117pp; English.  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II.  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 79; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 6,7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ASKNFHLQKNTIGTG 15  
 DB 240 ASKNFHLQKNTIGTG 254

Search completed: April 20, 2003, 13:06:11  
 Job time : 18.1974 secs



GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524d-131

Perfect score: 76

Sequence: 1 SRAEVSIVHNGAKP 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID2/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	100.0	15	17	AA97918
2	76	100.0	15	19	AA97918
3	76	100.0	15	19	AA97918
4	76	100.0	15	19	AA97918
5	76	100.0	15	19	AA97918
6	76	100.0	15	19	AA97918
7	76	100.0	15	19	AA97918
8	76	100.0	15	19	AA97918
9	76	100.0	15	19	AA97918
10	76	100.0	15	19	AA97918

11	72	94.7	14	18	AA97918
12	72	94.7	14	23	AA97918
13	72	94.7	14	23	AA97918
14	72	94.7	14	23	AA97918
15	63	82.9	134	18	AA97918
16	63	82.9	134	18	AA97918
17	61	80.3	514	17	AA97918
18	61	80.3	514	17	AA97918
19	59	77.6	507	22	AA97918
20	54	71.1	20	17	AA97918
21	50	65.8	15	17	AA97918
22	49	64.5	15	17	AA97918
23	43	56.6	415	21	AA97918
24	43	56.6	415	21	AA97918
25	43	56.6	415	21	AA97918
26	43	56.6	415	21	AA97918
27	42	55.3	1196	23	AA97918
28	42	55.3	1196	23	AA97918
29	40	52.6	578	21	AA97918
30	39	51.3	33	22	AA97918
31	39	51.3	127	22	AA97918
32	39	51.3	406	21	AA97918
33	39	51.3	430	21	AA97918
34	39	51.3	475	22	AA97918
35	38	50.0	116	21	AA97918
36	38	50.0	292	21	AA97918
37	38	50.0	300	21	AA97918
38	38	50.0	300	21	AA97918
39	38	50.0	391	23	AA97918
40	38	50.0	391	23	AA97918
41	38	50.0	392	22	AA97918
42	38	50.0	392	22	AA97918
43	38	50.0	398	22	AA97918
44	38	50.0	411	23	AA97918
45	38	50.0	433	21	AA97918

#### ALIGNMENTS

RESULT 1  
AA97918  
ID AA97918 standard; peptide; 15 AA.  
XX  
AC AA97918;  
XX  
DT 16-AUG-1996 (first entry)  
XX  
DE Japan cedar pollen mature allergen Cry j II amino acids 236-250.  
XX  
XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
KW Sugl pollinosis; diagnosis; treatment.  
XX  
OS Cryptomeria japonica.  
XX  
PN JP08047392-A.  
XX  
PD 20-FEB-1996.  
XX  
PF 07-NOV-1994; 94JP-0297840.  
XX  
PR 26-MAY-1994; 94JP-0134868.  
XX  
PR 05-NOV-1993; 93JP-0276773.  
XX  
PA (MEIP) MEIJI MILK PROD CO LTD.  
XX  
XX WPI; 1996-166249/17.  
XX  
XX Japan cedar pollen allergen Cry j II epitope - comprises at least  
XX part of specified 460 amino acid protein  
XX  
XX Claim 8; Fig 4; 17pp; Japanese.

Japanese cedar pol  
Immune tolerance a  
Multi-epitope pep  
Chamaecyparis obt  
Japanese cypress p  
Juniperus ashei N  
T-cell epitope pep  
Japan cedar pollen  
T-cell epitope pep  
Japan cedar pollen  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Herbicidally activ  
Arabidopsis thalia  
Human transporter  
Chlamydia pneumoni  
Haemophilus influe  
Human polypeptide  
Human polypeptide  
Arabidopsis thalia  
Arabidopsis thalia  
Staphylococcus aur  
Staphylococcus aur  
Staphylococcus aur  
S. epidermidis ope  
S. epidermidis ope  
Staphylococcus epi  
Arabidopsis thalia

CC AAR97871-R97960 are overlapping peptides used for the epitope mapping  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.

XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 17; Length 15;  
 Best Local Similarity 100.0%; Pred. NO. 1.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRAEVSIVHNGAKF 15  
 Db 1 SRAEVSIVHNGAKF 15

#### RESULT 2

AAW57765  
 ID AAW57765 standard; peptide; 15 AA.

XX  
 AC AAW57765;

XX  
 DT 17-SEP-1998 (first entry)

XX  
 DE Residues 236-250 of Cry j 2.

XX  
 KW Cry j 2; Japanese cedar pollen antigen; allergy; immunotherapy;

XX  
 KW HLA class II molecule.

XX  
 OS Cryptomeria japonica.

XX  
 PN WO9820902-A1.

XX  
 PD 22-MAY-1998.

XX  
 PF 12-NOV-1997; 97WO-JP04129.

XX  
 PR 13-NOV-1996; 96JP-0302053.

XX  
 PA (MEIP) MEIJI MILK PROD CO LTD.

XX  
 PI Dairiki K, Kino K, Kume A, Sone T;

XX  
 DR WPI; 1998-297617/26.

XX  
 PT Peptides derived from Japanese cedar pollen antigens are  
 PT immunotherapeutic agents - useful for allergy treatment and typing  
 PT HLA class II molecules in allergy sufferers

XX  
 PS Claim 12; Page 31; 50pp; Japanese.

XX  
 CC This sequence represents residues 236-250 of the Cry j 2 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.

XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 19; Length 15;  
 Best Local Similarity 100.0%; Pred. NO. 1.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRAEVSIVHNGAKF 15  
 Db 1 SRAEVSIVHNGAKF 15

#### RESULT 3

AAK69791  
 ID AAK69791 standard; Protein; 460 AA.

XX  
 AC AAK69791;

XX  
 DT 27-SEP-1995 (first entry)

XX  
 DE Japonicum allergen residues 55-514.

XX  
 KW Japonicum allergen; residues 55-514; induced histamine release;

XX  
 KW antiallergic peptide; IgE cross-linking inhibition.

XX  
 OS Japonicum sp.

XX  
 PN WO9502412-A.

XX  
 PD 26-JAN-1995.

XX  
 PF 15-JUL-1994; 94WO-JP01164.

XX  
 PR 16-JUL-1993; 93JP-0177008.

XX  
 PR 01-SEP-1993; 93JP-0217725.

XX  
 PR 07-APR-1994; 94JP-0069336.

XX  
 PA (MEIP) MEIJI MILK PROD CO LTD.

XX  
 PI Kino K, Kohno Y, Komiyama N, Sone T;

XX  
 DR WPI; 1995-067159/09.

XX  
 DR N-PSDB; AAK694044.

XX  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen

XX  
 PT with IgE antibody

XX  
 PS Disclosure; Pages 26-27; 46pp; Japanese.

XX  
 CC AAQ84044 encodes AAK69791 Japonicum allergen residues 55-514, from

XX  
 CC which the antiallergic peptides AAK69845-K69809 were derived.

XX  
 CC The peptides ability to inhibit the cross-linking of an allergen,  
 CC to an IgE antibody can be used in the prevention and treatment of  
 CC allergic diseases.

XX  
 SQ Sequence 460 AA;

OY 1 SRAEVSIVHNGAKF 15  
 Db 236 SRAEVSIVHNGAKF 250

XX  
 CC Query Match 100.0%; Score 76; DB 16; Length 460;  
 CC Best Local Similarity 100.0%; Pred. NO. 7.1e-06;  
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
 ID AAK53690 standard; Protein; 514 AA.

XX  
 AC AAK53690;

XX  
 DT 01-FEB-1995 (first entry)

XX  
 DE Japanese cedar pollen allergen Cry j II.

XX  
 KW Cedar pollinosis; diagnostic.

XX  
 OS Cryptomeria japonica.

PN W09411512-A.  
 XX 26-MAY-1994.  
 PD  
 XX  
 PF 12-NOV-1993; 93WO-US11000.  
 XX  
 PR 12-NOV-1992; 92US-0975179.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Brauer A, Kuo M, Pollock J, Yeung S;  
 XX  
 DR WPI; 1994-183513/22.  
 DR N-PSDB; AAQ66048.  
 XX  
 PT Allergic Cry j II protein and fragments from Japanese cedar  
 PT pollen - used to diagnose, treat and prevent Japanese cedar  
 XX pollinosis  
 XX  
 PS Claim 2; Fig 4; 89pp; English.  
 XX  
 CC The sequence is of a Japanese cedar pollen allergen Cry j  
 CC II. The protein and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also AAR53692-6.  
 CC  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 76; DB 15; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SRAEVSIVHNGAKF 15  
 DB 290 SRAEVSIVHNGAKF 304  
 RESULT 5  
 ID AAR74333 standard; Protein; 514 AA.  
 XX  
 AC AAR74333;  
 XX  
 DT 01-NOV-1995 (first entry)  
 XX  
 DE Japanese cedar pollen allergen.  
 XX  
 KW Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;  
 KW desensitizer.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN EP655500-A.  
 XX  
 PD 31-MAY-1995.  
 XX  
 PF 03-NOV-1994; 94EP-0308117.  
 XX  
 PR 27-DEC-1993; 93JP-0346814.  
 PR 05-NOV-1993; 93JP-029151.  
 PR 20-DEC-1993; 93JP-0344596.  
 XX  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI Kurimoto M, Namba M, Torigoe K;  
 XX  
 DR WPI; 1995-195588/26.  
 DR N-PSDB; AAQ90156.  
 XX  
 PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it,  
 PT useful for treatment and diagnosis of cedar pollen allergy  
 XX

PS Claim 5; Page 26-28; 41pp; English.  
 XX  
 CC The gene encoding an allergen of Japanese cedar pollen was isolated  
 CC by PCR amplification using primers based on portions of the allergen  
 CC protein. The gene was used for recombinant allergen production in  
 CC E. coli (vector plasmid pKK-223-3).  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 76; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SRAEVSIVHNGAKF 15  
 DB 290 SRAEVSIVHNGAKF 304  
 RESULT 6  
 ID AAR69792 standard; Protein; 514 AA.  
 XX  
 AC AAR69792;  
 XX  
 DT 27-SEP-1995 (first entry)  
 XX  
 DE Japonicum allergen.  
 XX  
 KW Japonicum allergen; induced histamine release; antiallergic peptide;  
 KW IGE cross-linking inhibition.  
 XX  
 OS Japonicum sp.  
 XX  
 PN W09502412-A.  
 XX  
 PD 26-JAN-1995.  
 XX  
 PF 15-JUL-1994; 94WO-JP01164.  
 XX  
 PR 16-JUL-1993; 93JP-0177008.  
 PR 01-SEP-1993; 93JP-0217725.  
 PR 07-APR-1994; 94JP-0069336.  
 XX  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX  
 PI Kino K, Kohno Y, Komiya N, Some T;  
 XX  
 DR WPI; 1995-067159/09.  
 DR N-PSDB; AAQ84045; AAQ84046.  
 XX  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 PT with IGE antibody  
 XX  
 PS Example 3; Pages 27-28; 46pp; Japanese.  
 XX  
 CC AAQ84045 encodes AAR69792 Japonicum allergen, from which the  
 CC antiallergic peptides AAR69845-R69809 were derived. The peptides ability  
 CC to inhibit the cross-linking of an allergen, to an IGE antibody can be  
 CC used in the prevention and treatment of allergic diseases.  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 76; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SRAEVSIVHNGAKF 15  
 DB 290 SRAEVSIVHNGAKF 304  
 RESULT 7  
 AAR93599

ID AAR93599 standard; Protein; 514 AA.  
 XX AAR93599;  
 AC  
 XX 16-AUG-1996 (first entry)  
 DT  
 XX Japan cedar pollen Cry j II allergen.  
 DE  
 XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
 XX Sugi pollinosis; diagnosis; treatment.  
 KW  
 XX Cryptomeria japonica.  
 OS  
 XX Key Location/Qualifiers  
 FH Peptide\* 1..54  
 FT /\*label= sig\_peptide  
 FT Protein 55..514  
 FT /\*label= mat\_protein  
 XX  
 XX JP08047392-A.  
 PN  
 XX 20-FEB-1996.  
 PD  
 XX 07-NOV-1994; 94JP-0297840.  
 PF  
 XX 26-MAY-1994; 94JP-0134868.  
 PR  
 XX 05-NOV-1993; 93JP-0276773.  
 XX  
 XX (MEIP ) MEIJI MILK PROD CO LTD.  
 PA  
 XX WPI; 1996-166249/17.  
 DR  
 XX N-PSDB; AAT18102.  
 XX  
 XX Japan cedar pollen allergen Cry j II epitope - comprises at least  
 PT part of specified 460 amino acid protein  
 XX  
 XX Claim 1; Page 10-11; 17pp; Japanese.  
 PS  
 XX AAR93599 is a Japan cedar pollen Cry j II allergen which is useful  
 CC in the diagnosis, prevention and treatment of Sugi pollinosis,  
 CC the allergic reaction to Japan cedar pollen. Significant regions of  
 CC the allergen were identified using overlapping peptides of the full  
 CC epitope derived from a Cry j II antigen-specific T cell line  
 CC (see AAR97871-R97860). Amino acids 66-80 (AAR97884) and 166-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.  
 XX  
 XX Sequence 514 AA;  
 SQ  
 Query Match 100.0%; Score 76; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SRAEVSIVHNGAKF 15  
 DB 290 SRAEVSIVHNGAKF 304  
 RESULT 8  
 ID AAR81586 standard; Protein; 514 AA.  
 AC AAR81586;  
 XX  
 XX 24-MAY-1996 (first entry)  
 DT  
 XX Cedar pollen allergen A.  
 DE  
 XX Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KW antibody; pollinosis; therapy; immunotherapy.  
 KW  
 XX Cryptomeria japonica.  
 OS  
 XX

PN EP700929-A2.  
 XX  
 XX 13-MAR-1996.  
 PD  
 XX  
 XX 08-SEP-1995; 95EP-0306295.  
 PF  
 XX 14-JUL-1995; 95JP-0200221.  
 PR 10-SEP-1994; 94JP-0242137.  
 PR 14-JUL-1995; 95JP-0200204.  
 XX  
 XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA  
 XX Hino K, Saito S, Taniguchi Y;  
 PI  
 XX WPI; 1996-140976/15.  
 DR  
 XX  
 XX New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PT used for treating cedar pollinosis  
 XX  
 XX Claim 5; Page 29-30; 36pp; English.  
 PS  
 XX Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 XX  
 XX Sequence 514 AA;  
 SQ  
 Query Match 100.0%; Score 76; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SRAEVSIVHNGAKF 15  
 DB 290 SRAEVSIVHNGAKF 304  
 RESULT 9  
 ID AAY25666 standard; protein; 514 AA.  
 AC AAY25666;  
 XX  
 XX 30-SEP-1999 (first entry)  
 DT  
 XX Japanese cedar allergen 1076242 Cry j II precursor protein fragment.  
 DE  
 XX Major histocompatibility complex, class II; desensitising; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 KW  
 XX Cedrus sp.  
 OS  
 XX WO9934826-A1.  
 PN  
 XX 15-JUL-1999.  
 PD  
 XX 11-JAN-1999; 99WO-GB00080.  
 PF  
 XX 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 PR  
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PA  
 XX Kay AB, Larche M;  
 PI



XX WP1; 1999-458255/38.  
 CC Desensitizing patients to polypeptide allergens  
 PT Example 6; Page 74; 117pp; English.  
 PS  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076242 Cry j II  
 CC precursor.  
 CC  
 SQ Sequence 514 AA;

Query Match 100.0%; Score 76; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRAEVSYPVHNGAKF 15  
 DB 290 SRAEVSYPVHNGAKF 304

## RESULT 10

AAV25667  
 ID AAV25667 standard; protein; 514 AA.

AC AAV25667;  
 DT 30-SEP-1999 (first entry)  
 XX

DE Japanese cedar allergen 1076241 Cry j II protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 XX  
 PN WO9934826-A1.  
 XX  
 PD 15-JUL-1999.  
 XX  
 PF 11-JAN-1999; 99WO-GB00080.  
 XX  
 PR 21-SEP-1998; 98GB-0020474.  
 XX  
 PR 09-JAN-1998; 98GB-0000445.  
 XX

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX

PI Kay AB, Larche M;  
 XX

DR WP1; 1999-458255/38.  
 XX

PT Desensitizing patients to polypeptide allergens  
 XX

PS Example 6; Page 74; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II.  
 CC  
 SQ Sequence 514 AA;

Query Match 100.0%; Score 76; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRAEVSYPVHNGAKF 15  
 DB 290 SRAEVSYPVHNGAKF 304

## RESULT 11

AAW12539  
 ID AAW12539 standard; peptide; 14 AA.

AC AAW12539;  
 DT 30-APR-1997 (first entry)  
 XX

DE Japanese cedar pollen allergen-derived peptide 8.  
 XX

KW Japanese cedar; sugi; pollen; pollinosis; pollenosis; allergen;  
 KW allergy; T-cell epitope; TCR; T-cell receptor; activation;  
 KW immune tolerance; Cryj.  
 XX

OS Cryptomeria japonica.  
 XX  
 OS Synthetic.  
 XX

PN JP08333391-A.  
 XX

PD 17-DEC-1996.  
 XX

PF 18-JUL-1995; 95UP-0181438.  
 XX

PR 07-APR-1995; 95UP-0082519.  
 XX

PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PA (SANY ) SANKYO CO LTD.  
 XX

DR WP1; 1997-095487/09.  
 XX

PT Peptide allergen derived from Japanese cedar pollen - causes T cell  
 XX response specific to cedar pollen, for treatment of pollenosis  
 PS Claim 11; Page 19; 21pp; Japanese.  
 XX

CC The present sequence is one of 24 claimed peptides which were  
 CC synthesised based on Japanese cedar pollen sequences. This peptide  
 CC was shown to induce tolerance in a BALB/c mouse. The peptide  
 CC produces little or no anaphylaxis.  
 XX

SQ Sequence 14 AA;

Query Match 94.7%; Score 72; DB 18; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RAEVSYVHNGAKF 15  
 |||||  
 DB 1 RAEVSYVHNGAKF 14

RESULT 12  
 AAU79882  
 ID AAU79882 standard; Peptide; 14 AA.

AC AAU79882;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Immune tolerance associated peptide #3.  
 XX  
 KM Immune tolerance; transgenic plant; T-cell epitope;  
 KM allergic protein.  
 XX  
 OS Cryptomeria japonica.  
 OS  
 PN JP2002085068-A.  
 XX  
 PD 26-MAR-2002.  
 XX  
 PF 08-SEP-2000; 2000JP-0272863.  
 XX  
 PR 08-SEP-2000; 2000JP-0272863.  
 XX  
 PA (SUMO) SUMITOMO CHEM CO LTD.  
 XX  
 DR WPI; 2002-378273/41.

XX  
 PT Induction of immune tolerance by oral administration of an effective  
 PT amount of transformed plant comprising a DNA coding for a polypeptide  
 PT having an amino acid sequence of T-cell epitope of an allergic protein  
 PT  
 PT  
 PS Example 4; Page 17; 19pp; Japanese.  
 XX  
 CC The invention describes a method for induction of immune tolerance  
 CC in animals, including humans. The treatment involves by oral  
 CC administration of an effective amount of transformed plant comprising  
 CC a DNA coding for a polypeptide having an amino acid sequence of T-cell  
 CC epitope of an allergic protein. The introduced protein is not degraded  
 CC by digestive enzymes in the host. This sequence represents a peptide  
 CC associated with the method of inducing immune tolerance discussed in  
 CC the invention.  
 CC  
 SQ Sequence 14 AA;

Query Match 94.7%; Score 72; DB 23; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RAEVSYVHNGAKF 15  
 |||||  
 DB 1 RAEVSYVHNGAKF 14

RESULT 13  
 AAW27371  
 ID AAW27371 standard; peptide; 134 AA.

AC AAW27371;  
 XX  
 DT 24-MAR-1998 (first entry)  
 XX  
 DE Multi-epitope peptide used as immunotherapeutic agent #3.  
 XX  
 KM Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KM T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 XX

XX  
 OS Synthetic.  
 XX  
 PN WO9732600-A1.  
 XX  
 PD 12-SEP-1997.  
 XX

PF 10-MAR-1997; 97WO-JP00740.  
 XX  
 PR 10-MAR-1996; 96JP-0080702.  
 XX

PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX

PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 XX  
 DR WPI, 1997-470495/43.  
 XX

PT Peptide immuno:therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 XX

PS Claim 6; Page 32; 58pp; Japanese.  
 XX

CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 XX  
 SQ Sequence 134 AA;

Query Match 94.7%; Score 72; DB 18; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RAEVSYVHNGAKF 15  
 |||||  
 DB 103 RAEVSYVHNGAKF 116

RESULT 14  
 AAW04346  
 ID AAW04346 standard; Protein; 514 AA.

AC AAW04346;  
 XX  
 DT 28-NOV-1996 (first entry)  
 XX  
 DE Chamaecyparis obtusa pollen allergen Cha o II.  
 XX  
 KM Pollen allergen; Cha o II; T-cell epitope; prevention; treatment;  
 KM pollinosis.  
 XX  
 OS Chamaecyparis obtusa.  
 XX  
 PN JP08176192-A.  
 XX  
 PD 09-JUL-1996.  
 XX  
 PF 21-DEC-1994; 94JP-0335089.  
 XX  
 PR 21-DEC-1994; 94JP-0335089.  
 XX

PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX

DR WPI; 1996-368225/37.  
 DR N-PSDB; AAT38521.  
 XX

PT DNA encoding chamaecyparis obtusa pollen allergen - T cell  
 PT epitope(s) of which are useful in development of preventative and  
 PT treating agent for C. obtusa pollen pollinosis

PS Claim 14; Pages 14-15; 17pp; Japanese.

CC The present sequence is the C. obtusa pollen allergen  
 CC Cha o II, the T-cell epitopes of which can be used in the  
 CC development of a preventive and treating agent for C. obtusa  
 CC pollen pollinosis. C. obtusa pollen (2.4 kg) was degassed with  
 CC diethyl ether, and dried at room temp. overnight. Cha o II was  
 CC sepd. from it and purified. RNA was extracted from C. obtusa  
 CC pollen, and mRNA and cDNA derived.

SQ Sequence 514 AA;

Query Match 82.9%; Score 63; DB 19; Length 514;

Best Local Similarity 86.7%; Pred. No. 0.0023;

Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SRAEVSIVHNGAKF 15

DB 290 SRAEVSIVHNGAKF 304

RESULT 15

AAW42122

ID AAW42122 standard; protein; 514 AA.

AC AAW42122;

DT 16-JUN-1998 (first entry)

DE Japanese cypress pollen antigen Cha2.

KW Japanese cypress pollen; antigen; T-cell epitope; Cha2; Cha2;  
 KW diagnosis; allergy; spring tree pollen disease; pollinosis.

OS Chamaecyparis obtusa.

PN WO9747648-A1.

PD 18-DEC-1997.

PF 12-JUN-1997; 97WO-JP02031.

PK 14-JUN-1996; 96JP-0153527.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kino K;

DR WPI, 1998-052242/05.

PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
 PT Cha2 and Cha2 - used for diagnosis and treatment of spring tree  
 PT pollen disease

PS Example 1; Page 15-18; 71pp; Japanese.

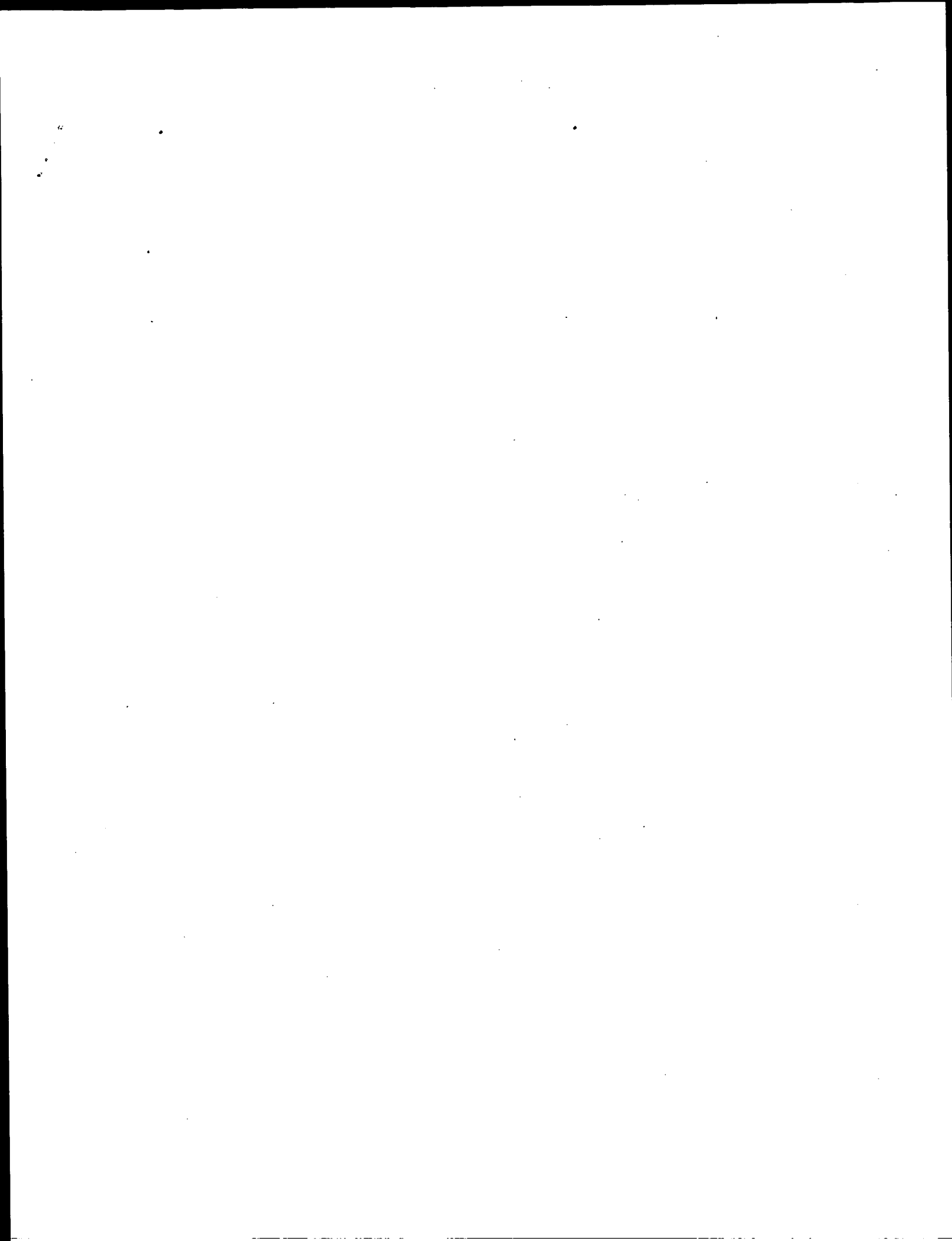
CC The present sequence represents Japanese cypress pollen antigen Cha2.  
 CC The present invention describes peptides which correspond to the T-cell  
 CC epitope sites on Japanese cypress pollen antigens Cha2 and Cha2. The  
 CC peptides can be used as a reagent for the diagnosis of allergy to  
 CC Japanese cypress pollen, and as an antigen in the treatment and  
 CC prevention of spring tree pollen disease in which the pollinosis  
 CC involves reactivity to Japanese cypress pollen.

SQ Sequence 514 AA;

QY 1 SRAEVSIVHNGAKF 15  
 DB 290 SRAEVSIVHNGAKF 304

Search completed: April 20, 2003, 13:06:12  
 Job time: 19.1974 secs

Query Match 82.9%; Score 63; DB 19; Length 514;  
 Best Local Similarity 86.7%; Pred. No. 0.0023;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds

(without alignments)  
109,838 Million cell updates/sec

Title: US-09-142-524d-152

Perfect score: 67

Sequence: 1 LSDSLKLTGSKIAS 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : \*

1: /SID2/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	100.0	15	17	AA897939
2	67	100.0	15	19	AA857770
3	67	100.0	460	16	AA859791
4	67	100.0	514	15	AA853630
5	67	100.0	514	16	AA874333
6	67	100.0	514	16	AA869792
7	67	100.0	514	17	AA893599
8	67	100.0	514	17	AA815866
9	67	100.0	514	20	AA825867
10	65	97.0	514	20	AA825666

11	59	88.1	17	17	AA815883	Cedar pollen aller
12	54	80.6	20	19	AAW42197	T-cell epitope pep
13	54	80.6	514	17	AAW04346	Chamaecyparis obtu
14	54	80.6	514	19	AAW42122	Japanese cypress p
15	53	79.1	17	19	AAW80348	Sugi allergen prot
16	53	79.1	80	18	AAW27369	Multi-epitope pep
17	51	76.1	453	22	AAW51693	Juniperus ashei N-
18	51	76.1	507	22	AAW51691	Juniperus ashei Ju
19	49	73.1	11	17	AAW81577	Cedar pollen aller
20	49	73.1	12	17	AAW81594	Cedar pollen aller
21	45	67.2	15	17	AAW14301	Japanese cedar pol
22	45	67.2	15	17	AAW97940	Japan cedar pollen
23	44	65.7	15	17	AAW97938	Japan cedar pollen
24	44	65.7	15	19	AAW57768	Residues 336-350 o
25	44	65.7	35	19	AAW80342	Sugi allergen prot
26	44	65.7	35	19	AAW80344	Sugi allergen prot
27	44	65.7	152	19	AAW38539	S. pneumoniae sul
28	44	65.7	242	18	AAW11194	S. pneumoniae sul
29	44	65.7	242	21	AAW81732	Streptococcus pneu
30	44	65.7	299	22	AAW37365	Staphylococcus aur
31	44	65.7	299	22	AAW37365	Staphylococcus aur
32	43	64.2	47	19	AAW80356	Sugi allergen prot
33	43	64.2	11	18	AAW14302	Japanese cedar pol
34	41	61.2	12	18	AAW14303	Japanese cedar pol
35	41	61.2	12	22	AAW69109	Cryptomeria japoni
36	41	61.2	33	19	AAW80340	Sugi allergen prot
37	41	61.2	33	19	AAW80339	Sugi allergen prot
38	41	61.2	33	19	AAW80341	Sugi allergen prot
39	41	61.2	35	19	AAW80343	Sugi allergen prot
40	41	61.2	47	19	AAW80350	Sugi allergen prot
41	41	61.2	47	19	AAW80352	Sugi allergen prot
42	41	61.2	47	19	AAW80352	Sugi allergen prot
43	41	61.2	47	19	AAW80353	Sugi allergen prot
44	41	61.2	47	19	AAW80357	Sugi allergen prot
45	41	61.2	51	21	AAW3877	Artificial sequenc

## ALIGNMENTS

RESULT 1  
AA897939  
ID AA897939 standard; peptide; 15 AA.  
XX AA897939;  
AC  
XX  
DT 16-AUG-1996 (first entry)  
XX  
DE Japan cedar pollen mature allergen Cry j II amino acids 341-355.  
XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
KW Sugi pollinosis; diagnosis; treatment.  
XX  
OS Cryptomeria japonica.  
XX  
PN JP08047392-A.  
XX  
PD 20-FEB-1996.  
XX  
PF 07-NOV-1994; 94JP-0297840.  
XX  
PR 26-MAY-1994; 94JP-0134868.  
XX  
PS 05-NOV-1993; 93JP-0276773.  
XX  
PA (MEIP) MEIJI MILK PROD CO LTD.  
XX  
DR WPI, 1996-166249/17.  
XX  
PT Japan cedar pollen allergen Cry j II epitope - comprises at least  
PT part of specified 460 amino acid protein  
PS Claim 8; Fig 5; 17pp; Japanese.  
XX

CC AAR97871-R97960 are overlapping peptides used for the epitope mapping  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.

XX Sequence 15 AA;

Query Match 100.0%; Score 67; DB 17; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LSDISLKTSGKIAS 15  
 |||||  
 Db 1 LSDISLKTSGKIAS 15

#### RESULT 2

ID AAW57770 standard; peptide; 15 AA.

XX AAW57770;

DT 17-SEP-1998 (first entry)

DE Residues 341-355 of Cry j 2.

XX Cry j 2; Japanese cedar pollen antigen; allergy; immunotherapy;

KW HLA class II molecule.

OS Cryptomeria japonica.

FN WO9820902-A1.

PD 22-MAY-1998.

PF 12-NOV-1997; 97WO-JP04129.

PR 13-NOV-1996; 96JP-0302053.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kino K, Kume A, Sone T;

DR WPI, 1998-297617/26.

PT Peptides derived from Japanese cedar pollen antigens are  
 PT immunotherapeutic agents - useful for allergy treatment and typing  
 PT HLA class II molecules in allergy sufferers

PS Claim 5; Page 33; 50pp; Japanese.

XX This sequence represents residues 341-355 of the Cry j 2 protein, and  
 XX is a peptide of the invention. The peptides are derived from Japanese  
 XX cedar pollen antigens, and are used as immunotherapeutic agents in the  
 XX treatment of allergy. The peptides can be used for identification and  
 XX typing of the particular HLA class II molecules in an allergy sufferer,  
 XX and also for peptide immunotherapy of an allergy. Using these peptides  
 XX the immunotherapy can be targeted more specifically to the requirements  
 XX of the individual patient, allowing more effective treatment of an  
 XX allergy, including those patients for whom treatment with a conventional  
 XX immunotherapeutic agent is ineffective.

SO Sequence 15 AA;

Query Match 100.0%; Score 67; DB 19; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LSDISLKTSGKIAS 15  
 |||||  
 Db 1 LSDISLKTSGKIAS 15

#### RESULT 3

ID AAR69791 standard; Protein; 460 AA.

XX AAR69791;

DT 27-SEP-1995 (first entry)

DE Japonicum allergen residues 55-514.

XX Japonicum allergen; residues 55-514; induced histamine release;

KW antiallergic peptide; IgE cross-linking inhibition.

OS Japonicum sp.

PN WO9502412-A.

PD 26-JAN-1995.

PF 15-JUL-1994; 94WO-JP01164.

PR 16-JUL-1993; 93JP-0177008.

PR 01-SEP-1993; 93JP-0217725.

PR 07-APR-1994; 94JP-0069336.

PI (MEIP) MEIJI MILK PROD CO LTD.

DR WPI, 1995-067159/09.

DR N-PSDB; AAO84044.

PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 PT with IgE antibody

PS Disclosure; Pages 26-27; 46pp; Japanese.

XX AAO84044 encodes AAR69791 Japonicum allergen residues 55-514, from  
 XX which the antiallergic peptides AAR69845-R69809 were derived.  
 XX The peptides ability to inhibit the cross-linking of an allergen,  
 XX to an IgE antibody can be used in the prevention and treatment of  
 XX allergic diseases.

SO Sequence 460 AA;

Query Match 100.0%; Score 67; DB 16; Length 460;  
 Best Local Similarity 100.0%; Pred. No. 0.0007;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LSDISLKTSGKIAS 15  
 |||||  
 Db 341 LSDISLKTSGKIAS 355

#### RESULT 4

ID AAR53690 standard; Protein; 514 AA.

XX AAR53690;

DT 01-FEB-1995 (first entry)

DE Japanese cedar pollen allergen Cry j II.

XX Cedar pollinosis; diagnostic.

OS Cryptomeria japonica.

PN WO9411512-A.  
 XX 26-MAY-1994.  
 XX 12-NOV-1993; 93WO-US11000.  
 XX 12-NOV-1992; 92US-0975179.  
 XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PA Brauer A, Kuo M, Pollock J, Yeung S;  
 XX WPI: 1994-183513/22.  
 DR N-PSDB; AAQ66048.  
 XX Allergic Cry j II protein and fragments from Japanese cedar  
 PT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 XX Claim 2; Fig 4; 89pp; English.  
 CC The sequence is of a Japanese cedar pollen allergen Cry j  
 CC II. The protein and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also AAR53692-6.  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 67; DB 15; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00079;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LSDISLKTSGKIAS 15  
 Db 395 LSDISLKTSGKIAS 409  
 RESULT 5  
 AAR74333  
 ID AAR74333 standard; Protein; 514 AA.  
 XX AAR74333;  
 AC  
 DT 01-NOV-1995 (first entry)  
 XX  
 DE Japanese cedar pollen allergen.  
 XX Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;  
 KW desensitizer.  
 XX Cryptomeria japonica.  
 OS  
 XX EP65500-A.  
 PN  
 XX 31-MAY-1995.  
 PD  
 XX 03-NOV-1994; 94EP-0308117.  
 PF  
 XX 27-DEC-1993; 93JP-0346814.  
 PR  
 XX 05-NOV-1993; 93JP-0299151.  
 PR  
 XX 20-DEC-1993; 93JP-0344596.  
 XX  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI Kurimoto M, Namba M, Torigoe K;  
 XX  
 DR WPI, 1995-195588/26.  
 DR N-PSDB; AAQ90156.  
 XX  
 PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it,  
 PT useful for treatment and diagnosis of cedar pollen allergy  
 XX

PS Claim 5; Page 26-28; 41pp; English.  
 CC The gene encoding an allergen of Japanese cedar pollen was isolated  
 CC by PCR amplification using primers based on portions of the allergen  
 CC protein. The gene was used for recombinant allergen production in  
 CC E. coli (vector plasmid pKK-223-3).  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 67; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00079;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LSDISLKTSGKIAS 15  
 Db 395 LSDISLKTSGKIAS 409  
 RESULT 6  
 AAR69792  
 ID AAR69792 standard; Protein; 514 AA.  
 XX AAR69792;  
 AC  
 DT 27-SEP-1995 (first entry)  
 XX  
 DE Japonicum allergen.  
 XX Japonicum allergen; induced histamine release; antiallergic peptide;  
 KW Ige cross-linking inhibition.  
 XX Japonicum sp.  
 PN WO9502412-A.  
 PD 26-JAN-1995.  
 XX  
 XX 15-JUL-1994; 94WO-JP01164.  
 PF  
 XX 16-JUL-1993; 93JP-0177008.  
 PR  
 XX 01-SEP-1993; 93JP-0217725.  
 PR  
 XX 07-APR-1994; 94JP-0069336.  
 XX  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX  
 PI Kino K, Kohno Y, Komiyama N, Sone T;  
 XX  
 DR WPI, 1995-067159/09.  
 DR N-PSDB; AAQ84045, AAQ84046.  
 XX  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 PT with Ige antibody  
 XX  
 PS Example 3; Pages 27-28; 46pp; Japanese.  
 XX  
 CC AAQ84045 encodes AAR69792 Japonicum allergen, from which the  
 CC antiallergic peptides AAR69845-R69809 were derived. The peptides ability  
 CC to inhibit the cross-linking of an allergen, to an Ige antibody can be  
 CC used in the prevention and treatment of allergic diseases.  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 67; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00079;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LSDISLKTSGKIAS 15  
 Db 395 LSDISLKTSGKIAS 409  
 RESULT 7  
 AAR93599

ID AAR93599 standard; Protein; 514 AA.  
 XX AAR93599;  
 AC  
 XX 16-AUG-1996 (first entry)  
 DT  
 XX Japan cedar pollen Cry j II allergen.  
 DE  
 XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
 KM Sugi pollinosis; diagnosis; treatment.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..54  
 FT /\*label= sig\_peptide  
 FT Protein 55..514  
 FT /\*label= mat\_protein  
 XX  
 FN JP08047392-A.  
 XX  
 PD 20-FEB-1996.  
 XX  
 PF 07-NOV-1994; 94JP-0297840.  
 XX  
 PR 26-MAY-1994; 94JP-0134868.  
 PR 05-NOV-1993; 93JP-0276773.  
 XX  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 XX  
 DR WPI; 1996-166249/17.  
 DR N-PSDB; AAT18102.  
 XX  
 PT Japan cedar pollen allergen Cry j II epitope - comprises at least  
 PI part of specified 460 amino acid protein  
 XX  
 PS Claim 1; Page 10-11; 17pp; Japanese.  
 XX  
 CC AAR93599 is a Japan cedar pollen Cry j II allergen which is useful  
 CC in the diagnosis, prevention and treatment of Sugi pollinosis,  
 CC the allergic reaction to Japan cedar pollen. Significant regions of  
 CC the allergen were identified using overlapping peptides of the full  
 CC epitope derived from a Cry j II antigen-specific T cell line  
 CC (see AAR97871-R97860). Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.  
 CC  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 67; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00079; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0;  
 OY 1 LSDISIKLTSGKIAS 15  
 DB 395 LSDISIKLTSGKIAS 409  
 RESULT 8  
 ID AAR81586 standard; Protein; 514 AA.  
 XX AAR81586;  
 AC  
 XX 24-MAY-1996 (first entry)  
 DT  
 XX Cedar pollen allergen A.  
 DE  
 XX Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KM antibody; pollinosis; therapy; immunotherapy.  
 XX  
 OS Cryptomeria japonica.  
 XX

PN EP700929-A2.  
 XX  
 PD 13-MAR-1996.  
 XX  
 PF 08-SEP-1995; 95EP-0306295.  
 XX  
 PR 14-JUL-1995; 95JP-0200221.  
 PR 10-SEP-1994; 94JP-0242137.  
 PR 14-JUL-1995; 95JP-0200204.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI Hino K, Saito S, Taniguchi Y;  
 XX  
 DR WPI; 1996-140976/15.  
 XX  
 PT New peptide(s) derived from cedar pollen allergens - activate  
 PI allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 XX used for treating cedar pollinosis  
 XX  
 PS Claim 5; Page 29-30; 36pp; English.  
 XX  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 CC  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 67; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00079; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0;  
 OY 1 LSDISIKLTSGKIAS 15  
 DB 395 LSDISIKLTSGKIAS 409  
 RESULT 9  
 ID AAY25667 standard; Protein; 514 AA.  
 XX AAY25667;  
 AC  
 XX 30-SEP-1999 (first entry)  
 DT  
 XX Japanese cedar allergen 1076241 Cry j II protein fragment.  
 DE  
 XX Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 XX  
 PN WO9934826-A1.  
 XX  
 PD 15-JUL-1999.  
 XX  
 PF 11-JAN-1999; 99WO-GB00080.  
 XX  
 PR 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX  
 PI Kay AB, Larche M;



XX WPI; 1999-458255/38.  
 XX Desensitizing patients to polypeptide allergens  
 PT  
 XX  
 PS  
 XX  
 Example 6; Page 74; 117pp; English.

CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076242 Cry j II.  
 XX  
 SQ Sequence 514 AA;

Query Match 100.0%; Score 67; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00079;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LSDISLKTSGKIAS 15  
 |||||  
 Db 395 LSDISLKTSGKIAS 409

RESULT 10  
 AAY25666  
 ID AAY25666 standard; protein; 514 AA.  
 XX  
 AC AAY25666;

DT 30-SEP-1999 (first entry)  
 XX

DE Japanese cedar allergen 1076242 Cry j II precursor protein fragment.  
 XX

KM Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX

OS Cedrus sp.  
 XX

EN WO9934826-A1.  
 XX

PD 15-JUL-1999.  
 XX

PF 11-JAN-1999; 99WO-GB00080.  
 XX

PR 21-SEP-1998; 98GB-0020474.  
 XX

PR 09-JAN-1998; 98GB-0000445.  
 XX

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX

PI Kay AB, Larche M;  
 XX

DR WPI; 1999-458255/38.  
 XX

PT Desensitizing patients to polypeptide allergens  
 XX

PS Example 6; Page 74; 117pp; English.  
 XX

CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076242 Cry j II  
 precursor.  
 XX

SQ Sequence 514 AA;

Query Match 97.0%; Score 65; DB 20; Length 514;  
 Best Local Similarity 93.3%; Pred. No. 0.0018;  
 Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LSDISLKTSGKIAS 15  
 |||||  
 Db 395 LSDISLKTSGKIAS 409

RESULT 11  
 AAR81583  
 ID AAR81583 standard; Peptide; 17 AA.  
 XX  
 AC AAR81583;

DT 24-MAY-1996 (first entry)  
 XX

DE Cedar pollen allergen peptide 11 (T-cell epitope).  
 XX

KM Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;  
 KM IgE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.  
 XX

OS Synthetic.  
 XX

PN EP700929-A2.  
 XX

PD 13-MAR-1996.  
 XX

PF 08-SEP-1995; 95EP-0306295.  
 XX

PR 14-JUL-1995; 95JP-0200221.  
 XX

PR 10-SEP-1994; 94JP-0242137.  
 XX

PR 14-JUL-1995; 95JP-0200204.  
 XX

PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX

PI Hino K, Saito S, Taniguchi Y;  
 XX

DR WPI; 1996-140976/15.  
 XX

PT New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PT used for treating cedar pollinosis  
 XX

PS Claim 4; Page 29; 36pp; English.  
 XX

CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar

CC pollinosis, avoiding side-effects such as anaphylaxis.

XX Sequence 17 AA;

Query Match 88.1%; Score 59; DB 17; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.0049;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DISLKLTSKGIAS 15  
DB 1 DISLKLTSKGIAS 13

# RESULT 12

AAW42197  
ID AAW42197 standard; peptide; 20 AA.

AC AAW42197;

DT 16-JUN-1998 (first entry)

DE T-cell epitope peptide 77 from Japanese cypress pollen antigen Cha2.

KW Japanese cypress pollen; antigen; T-cell epitope; Chaol; Cha2;

KW diagnosis; allergy; spring tree pollen disease; pollinosis.

OS Chamaecyparis obtusa.

PN WO9747648-A1.

PD 18-DEC-1997.

PF 12-JUN-1997; 97WO-JP02031.

PR 14-JUN-1996; 96JP-0153527.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kino K;

DR WPI; 1998-052242/05.

PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
PT Chaol and Cha2 - used for diagnosis and treatment of spring tree  
PT pollen disease

PS Claim 2; Page 50; 71pp; Japanese.

CC The present sequence represents a T-cell epitope peptide from Japanese  
CC cypress pollen antigen Cha2. The present invention describes peptides  
CC which correspond to the T-cell epitope sites on Japanese cypress pollen  
CC antigens Chaol and Cha2. The peptides can be used as a reagent for the  
CC diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
CC the treatment and prevention of spring tree pollen disease in which the  
CC pollinosis involves reactivity to Japanese cypress pollen.

SQ Sequence 20 AA;

Query Match 80.6%; Score 54; DB 19; Length 20;  
Best Local Similarity 80.0%; Pred. No. 0.0049;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGIAS 15  
DB 5 LSNVSLKLTSKGPAS 19

# RESULT 13

AAW04346  
ID AAW04346 standard; Protein; 514 AA.

AC AAW04346;  
XX

DT 28-NOV-1996 (first entry)

XX Chamaecyparis obtusa pollen allergen Cha o II.

XX Pollen allergen; Cha o II; T-cell epitope; prevention; treatment;  
XX pollinosis.

OS Chamaecyparis obtusa.

PN JP08176192-A.

PD 09-JUL-1996.

PF 21-DEC-1994; 94JP-0335089.

PR 21-DEC-1994; 94JP-0335089.

PA (MEIP) MEIJI MILK PROD CO LTD.

DR WPI; 1996-36825/37.

DR N-PSDB; AAT38521.

PT DNA encoding chamaecyparis obtusa pollen allergen - T cell  
PT epitope(s) of which are useful in development of preventative and  
PT treating agent for C. obtusa pollen pollinosis

PS Claim 14; Pages 14-15; 17pp; Japanese.

CC The present sequence is the C. obtusa pollen allergen  
CC Cha o II, the T-cell epitopes of which can be used in the  
CC development of a preventative and treating agent for C. obtusa  
CC pollen pollinosis. C. obtusa pollen (2.4 Kg) was degassed with  
CC diethyl ether, and dried at room temp. overnight. Cha o II was  
CC seed, from it and purified. RNA was extracted from C. obtusa  
CC pollen, and mRNA and cDNA derived.

SQ Sequence 514 AA;

Query Match 80.6%; Score 54; DB 17; Length 514;  
Best Local Similarity 80.0%; Pred. No. 0.19;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGIAS 15  
DB 395 LSNVSLKLTSKGPAS 409

# RESULT 14

AAW42122  
ID AAW42122 standard; protein; 514 AA.

AC AAW42122;

DT 16-JUN-1998 (first entry)

DE Japanese cypress pollen antigen Cha2.

KW Japanese cypress pollen; antigen; T-cell epitope; Chaol; Cha2;

KW diagnosis; allergy; spring tree pollen disease; pollinosis.

OS Chamaecyparis obtusa.

PN WO9747648-A1.

PD 18-DEC-1997.

PF 12-JUN-1997; 97WO-JP02031.

PR 14-JUN-1996; 96JP-0153527.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kino K;

XX WPI; 1998-052242/05.  
 DR T-cell epitope peptide portion of Japanese cypress pollen antigens  
 XX Chaol and Chaol2 - used for diagnosis and treatment of spring tree  
 PT pollen disease  
 PT Example 1; Page 15-18, 71pp; Japanese.  
 XX  
 CC The present sequence represents Japanese cypress pollen antigen Chaol2.  
 CC The present invention describes peptides which correspond to the T-cell  
 CC epitope sites on Japanese cypress pollen antigens Chaol and Chaol2. The  
 CC peptides can be used as a reagent for the diagnosis of allergy to  
 CC Japanese cypress pollen, and as an antigen in the treatment and  
 CC prevention of spring tree pollen disease in which the pollinosis  
 CC involves reactivity to Japanese cypress pollen.  
 XX  
 SQ Sequence 514 AA;

Query Match Best Local Similarity 80.6%; Score 54; DB 19; Length 514;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LSDISLKTSGKIAS 15  
 ||:|||||  
 DB 395 LSNVSLKTSGKPDAS 409

## RESULT 15

AAW80348  
 ID AAW80348 standard; peptide; 17 AA.

AAW80348;  
 AC  
 XX

DT 11-JAN-1999 (first entry)

DE Sugi allergen protein Cryj2 derived epitope for T cells.

XX T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
 KW sugi-pollinosis; allergic reaction; pollen.  
 XX

OS Synthetic.  
 XX

PN JP10259198-A.  
 XX

PD 29-SEP-1998.  
 XX

PF 22-DEC-1997; 97JP-0353448.  
 XX

PR 24-DEC-1996; 96JP-0343441.  
 XX

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX (SANY) SANKYO CO LTD.  
 XX

DR WPI; 1998-577037/49.  
 XX

PT A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 XX

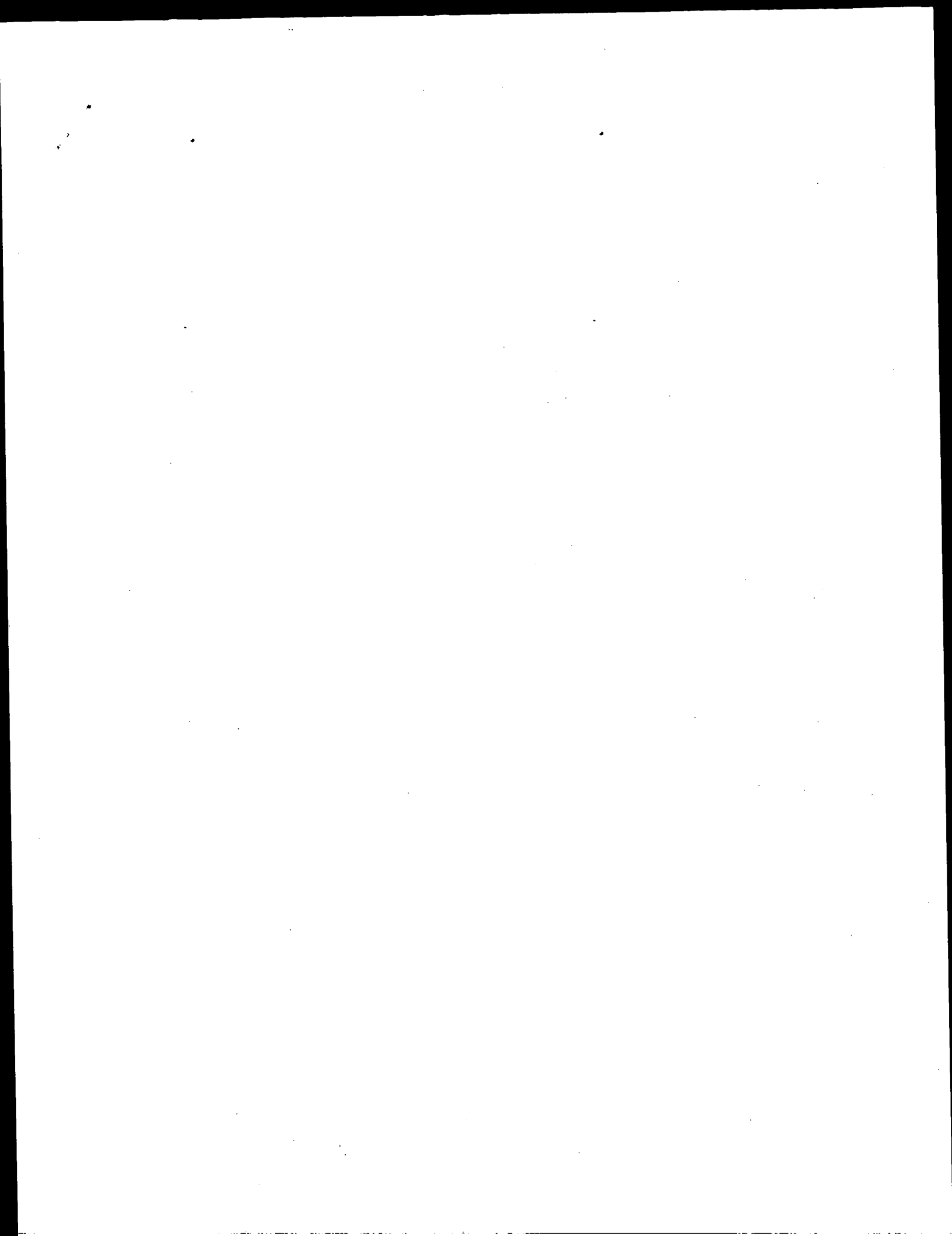
PS Claim 7; Page 18; 21pp; Japanese.  
 XX

CC AAW80339-58 represent epitopes for T cells, derived from the sugi  
 CC allergen proteins Cryj1 (AAW80339-44, AAW80350-53 and AAW80356-58) and  
 CC Cryj2 (AAW80345-49 and AAW80354-55). The peptides are useful for the  
 CC treatment of sugi-pollinosis, an allergic reaction of the body to  
 CC pollen.  
 XX

SQ Sequence 17 AA;

QY 4 ISLKTSGKIAS 15  
 |||||||  
 DB 1 ISLKTSGKIAS 12  
 Search completed: April 20, 2003, 13:06:13  
 Job time : 19.1974 secs

Query Match Best Local Similarity 79.1%; Score 53; DB 19; Length 17;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments) 237.215 Million cell updates/sec

Title: US-09-142-524D-28  
Perfect score: 81  
Sequence: 1 RPLMIIFSGNNMIKL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	100.0	374	2	JC2124
2	81	100.0	374	2	JC2123
3	58	71.6	397	2	S26211
4	57	70.4	542	2	T06728
5	56	69.1	455	2	T00856
6	56	69.1	463	2	T46165
7	55	67.9	449	2	S27098
8	54	66.7	398	2	B53240
9	54	66.7	398	2	B39099
10	53	65.4	392	2	D53240
11	53	65.4	472	2	T51456
12	52	64.2	396	2	A39099
13	52	64.2	397	2	C53240
14	52	64.2	397	2	C38099
15	52	64.2	434	2	S29612
16	51	63.0	438	2	S43335
17	51	63.0	459	2	G86278
18	50	61.7	274	2	T03932
19	50	61.7	431	2	P86179
20	49	60.5	404	2	T05556
21	49	60.5	501	2	T34741
22	48	59.3	397	2	E53240
23	47	58.0	385	2	A91006
24	47	58.0	385	2	B85850
25	46	56.8	398	2	A64981
26	46	56.8	398	2	T07058
27	45	55.6	404	2	S12209
28	45	55.6	251	2	C86248
29	45	55.6	374	2	T05240

## ALIGNMENTS

30	45	55.6	374	2	H85148	probable pectate 1
31	44	54.3	450	2	T09524	probable pectate 1
32	43	53.1	341	2	T47653	pectate lyase-like
33	43	53.1	418	2	T07701	pectate lyase (EC
34	42	53.1	839	2	T21207	hypothetical prote
35	42	51.9	344	2	D86141	protein T25K16.7 l
36	42	51.9	350	2	D84812	probable peroxidase
37	42	51.9	394	2	T49115	pectate lyase like
38	42	51.9	394	2	T49116	pectate lyase like
39	42	51.9	550	2	T40486	phosphoinositide-d
40	41	50.6	137	2	B75075	hypothetical prote
41	41	50.6	174	2	A86693	hypothetical prote
42	41	50.6	368	2	G86427	probable pectate 1
43	41	50.6	390	2	H86253	hypothetical prote
44	41	50.6	915	2	H82104	cation transport A
45	41	50.6	1466	2	T39557	vacuolar protein s

## RESULT 1

JC2124  
major allergen Cry j I precursor (clone PCCI-15) - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000  
C/Accession: JC2124  
R/Source, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan  
A/Reference number: JC2123; MUID:94189234; PMID:8135802  
A/Accession: JC2124  
A/Molecule type: mRNA  
A/Residues: 1-374 <SON>  
A/Cross-References: GB:D26545; NID:G493633; PIDN:BA05543.1; PID:G493634  
A/Experimental source: pollen  
A/Note: the authors described carbohydrate binding site for residue 279  
C/Superfamily: pectate lyase LAM59  
C/Keywords: glycoprotein; pollen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen Cry j I (clone PCCI-15) #status predicted <MAT>  
F:158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 81; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNMIKL 15  
DB 87 RPLMIIFSGNNMIKL 101

RESULT 2  
JC2123  
major allergen Cry j I precursor (clone PCCI-2-2) - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000  
C/Accession: JC2123; PC2065  
R/Source, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan  
A/Reference number: JC2123; MUID:94183234; PMID:8135802  
A/Accession: JC2123  
A/Molecule type: mRNA  
A/Residues: 1-374 <SON>  
A/Cross-References: GB:D26544; NID:G493631; PIDN:BA05542.1; PID:G493632  
A/Experimental source: pollen  
A/Accession: PC2065  
A/Molecule type: protein  
A/Residues: 22-53; 58-81; 219-232; 236-258; 299-307; 346-372 <SO2>  
A/Note: the authors described carbohydrate binding site for residue 279  
C/Superfamily: pectate lyase LAM59  
C/Keywords: glycoprotein; pollen

F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-374/Product: major allergen Cry I (clone PCCI-2-2) #status predicted <MA>  
 F:150,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 81; DB 2; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RPLWIFSGNNIKL 15  
 DB 87 RPLWIFSGNNIKL 101

## RESULT 3

S26211  
 pectate lyase (EC 4.2.2.2) - common tobacco  
 C/Species: Nicotiana tabacum (common tobacco)  
 C/Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 21-Jul-2000  
 C/Accession: S26211; S26212; S21933; S22753; S22754  
 R/Rogers, H.J.; Harvey, A.; Lonsdale, D.M.  
 Plant Mol. Biol. 20, 493-502, 1992

A/Title: Isolation and characterization of a tobacco gene with homology to pectate lyase  
 A/Reference number: S26211; MUID:93043039; PMID:1421152

A/Accession: S26211  
 A/Molecule type: DNA  
 A/Residues: 1-397 <ROG>  
 A/Cross-references: EMBL:X67158; NID:G19907; PIDN:CAA47630.1; PID:G19908  
 A/Accession: S26212  
 A/Molecule type: mRNA  
 A/Residues: 119-155; 'C', 157-188, 190, 'G', 191-199, 'D', 200, 'R', 203-248, 'N', 250-381 <RO2>  
 A/Cross-references: EMBL:X67159; NID:G19909; PIDN:CAA47631.1; PID:G3980174  
 A/Note: translation of the nucleotide sequence is not complete  
 R/Lonsdale, D.M.  
 submitted to the EMBL Data Library, July 1991

A/Reference number: S21933

A/Accession: S21933  
 A/Molecule type: DNA  
 A/Residues: 1-397 <LON>  
 A/Cross-references: EMBL:X61102; NID:G19981; PIDN:CAA3414.1; PID:G19982

C/Genetics:

A/Introns: 193/1; 293/2

C/Superfamily: pectate lyase LAT59

C/Keywords: carbon-oxygen lyase

Query Match 71.6%; Score 58; DB 2; Length 397;  
 Best Local Similarity 78.6%; Pred. No. 0.029;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 RPLWIFSGNNIKL 15  
 DB 112 RPLWIFSGNNIKL 125

## RESULT 4

T06728  
 pectate lyase (EC 4.2.2.2) F28P10.100 - Arabidopsis thaliana

N/Alternate names: protein F28P10.100

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999

C/Accession: T06728

R/Querier, F.; Choise, N.; Robert, C.; Brotier, P.; Winkler, P.; Catolico, L.; Artigou

submitted to the Protein Sequence Database, April 1999

A/Reference number: Z15793

A/Accession: T06728

A/Molecule type: DNA

A/Residues: 1-542 <QUB>

A/Cross-references: EMBL:AL049655

A/Experimental source: cultivar Columbia; BAC clone F28P10

C/Genetics:

A/Map position: 3

A/Introns: 46/2; 346/3; 413/2; 480/2

A/Note: F28P10.100

C/Superfamily: pectate lyase LAT59

C/Keywords: carbon-oxygen lyase

Query Match 70.4%; Score 57; DB 2; Length 542;  
 Best Local Similarity 71.4%; Pred. No. 0.061;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 RPLWIFSGNNIKL 15  
 DB 146 RPLWIFSGNNIKL 159

## RESULT 5

T00856  
 pectate lyase (EC 4.2.2.2) T20P6.14 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Feb-2001

C/Accession: T00856; A84440

R/Rounsley, S.D.; Lin, X.; Ketchum, R.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, March 1998

A/Description: Arabidopsis thaliana chromosome II BAC T20P6 genomic sequence.

A/Reference number: Z14206

A/Accession: T00856

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-455 <ROU>

A/Cross-references: EMBL:AC002521; NID:G2947056; PIDN:AAC05350.1; PID:G2947069

A/Experimental source: cultivar Columbia

R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

usens, D.; Niekman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: A84440

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-455 <STO>

A/Cross-references: GB:AE002093; NID:G2947069; PIDN:AAC05350.1; GSPDB:GND00139

C/Genetics:

A/Map position: 2

A/Introns: 66/2; 295/3; 376/3

A/Note: T20P6.14

C/Superfamily: pectate lyase LAT59

C/Keywords: carbon-oxygen lyase

Query Match 69.1%; Score 56; DB 2; Length 455;  
 Best Local Similarity 73.3%; Pred. No. 0.075;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 RPLWIFSGNNIKL 15  
 DB 166 RPLWIFSGNNIKL 180

T46165  
 pectate lyase-like protein - Arabidopsis thaliana

N/Alternate names: protein T46165

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 02-Sep-2000

C/Accession: T46165

R/Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, F

submitted to the Protein Sequence Database, December 1999

A/Reference number: Z23025

A/Accession: T46165

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-463 <NTA>

A/Experimental source: cultivar Columbia; BAC clone T4D2

C/Genetics:

A/Map position: 3

A:Introns: 27/2; 112/2; 182/1; 265/3; 329/3; 399/2  
 A:Note: T4D2.120  
 C:Superfamily: pectate lyase LAT59

Query Match 69.1%; Score 56; DB 2; Length 463;  
 Best Local Similarity 71.4%; Pred. No. 0.077;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLWIFSGNNMIXL 15  
 DB 120 PLWIFPSNNMIXL 133

## RESULT 7

pectate lyase (EC 4.2.2.2) LAT59 - tomato  
 N:Alternate names: protein P59  
 C:Species: Lycopersicon esculentum (tomato)  
 C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 20-Apr-2001  
 C:Accession: S27098; S08064  
 R:Griffith, I.J.; Yamaguchi, J.; Larrabell, S.K.; Ursin, V.M.; McCormick, S.  
 Plant Mol. Biol. 14, 17-28, 1990  
 A:Title: Molecular and genetic characterization of two pollen-expressed genes that have  
 A:Reference number: S08383; MUID:91322485; PMID:1983191  
 A:Accession: S27098  
 A:Molecule type: DNA  
 A:Residues: 1-448 <WIN>  
 A:Cross-references: EMBL:X15499; NID:919270; PIDN:CNA33523.1; PID:919271  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 67/2; 177/2  
 C:Superfamily: pectate lyase LAT59  
 C:Keywords: carbon-oxygen lyase

Query Match 67.9%; Score 55; DB 2; Length 449;  
 Best Local Similarity 71.4%; Pred. No. 0.11;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLWIFSGNNMIXL 15  
 DB 165 PLWIFKRGNNMIXL 178

## RESULT 8

B53240  
 allergen Amb a 1.2 precursor - common ragweed  
 C:Species: Ambrosia artemisiifolia (common ragweed)  
 C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 18-Sep-1998  
 C:Accession: B53240  
 R:Griffith, I.J.; Pollock, J.; Klapper, D.G.; Rogers, B.L.; Nault, A.K.  
 Int. Arch. Allergy Appl. Immunol. 96, 296-304, 1991  
 A:Title: Sequence polymorphism of Amb a I and Amb a II, the major allergens in Ambrosia  
 A:Reference number: A53240; MUID:92234570; PMID:1809687  
 A:Accession: B53240  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-398 <GRI>  
 A:Cross-references: GB:M0559  
 C:Superfamily: pectate lyase LAT59  
 C:Keywords: pollen

Query Match 66.7%; Score 54; DB 2; Length 398;  
 Best Local Similarity 73.3%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNMIXL 15  
 DB 115 RPLWIFKRGNNMIXL 129

RESULT 9  
 B39099  
 allergen Amb a 1.2 - common ragweed

C:Species: Ambrosia artemisiifolia (common ragweed)  
 C:Date: 27-Nov-1991 #sequence\_revision 03-Apr-1992 #text\_change 21-Jul-2000  
 C:Accession: B39099

R:Refnar, T.; Griffith, I.J.; Kuo, M.; Bond, J.F.; Rogers, B.L.; Klapper, D.G.  
 J. Biol. Chem. 266, 1229-1236, 1991  
 A:Title: Cloning of Amb a I (antigen E), the major allergen family of short ragweed  
 A:Reference number: A39099; MUID:91093235; PMID:1702434  
 A:Accession: B39099  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-398 <RAF>  
 A:Cross-references: GB:M0581; NID:9166436; PIDN:AAA32666.1; PID:9166437  
 C:Superfamily: pectate lyase LAT59  
 C:Keywords: pollen

Query Match 66.7%; Score 54; DB 2; Length 398;  
 Best Local Similarity 73.3%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNMIXL 15  
 DB 115 RPLWIFKRGNNMIXL 129

## RESULT 10

D53240  
 allergen Amb a 1.4 precursor - common ragweed  
 C:Species: Ambrosia artemisiifolia (common ragweed)  
 C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 29-Sep-1999  
 C:Accession: D53240  
 R:Griffith, I.J.; Pollock, J.; Klapper, D.G.; Rogers, B.L.; Nault, A.K.  
 Int. Arch. Allergy Appl. Immunol. 96, 296-304, 1991  
 A:Title: Sequence polymorphism of Amb a I and Amb a II, the major allergens in Ambrosia  
 A:Reference number: A53240; MUID:92234570; PMID:1809687  
 A:Accession: D53240  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-392 <GRI>  
 A:Cross-references: GB:M0562; NID:9166444; PIDN:AAA32670.1; PID:9166445  
 C:Superfamily: pectate lyase LAT59  
 C:Keywords: pollen

Query Match 65.4%; Score 53; DB 2; Length 392;  
 Best Local Similarity 66.7%; Pred. No. 0.21;  
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNMIXL 15  
 DB 114 RPLWIFKRGNNMIXL 128

## RESULT 11

T51456  
 pectate lyase-like protein - Arabidopsis thaliana  
 N:Alternate names: protein F2G14\_230  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 17-Nov-2000  
 C:Accession: T51456  
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; N  
 submitted to the Protein Sequence Database, August 2000  
 A:Reference number: Z25394  
 A:Accession: T51456  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-472 <SAT>  
 A:Cross-references: EMBL:AL391146  
 A:Experimental source: cultivar Columbia; BAC clone F2G14  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 65/2; 316/3; 398/3  
 A:Note: F2G14\_230  
 C:Superfamily: pectate lyase LAT59

Query Match 65.4%; Score 53; DB 2; Length 472;  
 Best Local Similarity 64.3%; Pred. No. 0.26;  
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 RPLMIIFSGNNIKL 15  
 ||||| :|||  
 Db 188 RPLMIIFKNDMVIRL 201

## RESULT 12

allergen Amb a 1.1 precursor - common ragweed  
 A:Accession: A39099  
 C:Species: Ambrosia artemisiifolia (common ragweed)  
 C>Date: 27-Nov-1991 #sequence\_revision 03-Apr-1992 #text\_change 29-Sep-1999  
 C:Accession: A39099; A60895; A53240  
 R:Reiner, T.; Griffith, I.J.; Kuo, M.; Bond, J.F.; Rogers, B.L.; Klapper, D.G.  
 J. Biol. Chem. 266, 1229-1236, 1991  
 A>Title: Cloning of Amb a I (antigen E), the major allergen family of short ragweed pollen  
 A:Reference number: A39099; PMID:91093235; PMID:1702434  
 A:Accession: A39099  
 A:Molecule type: mRNA  
 A:Residues: 1-396 <RA>  
 A:Cross-references: GB:M63116  
 R:Smith, J.J.; Olson, J.R.; Klapper, D.G.  
 Mol. Immunol. 25, 355-365, 1988  
 A>Title: Monoclonal antibodies to denatured ragweed pollen allergen Amb a I: characterization  
 A:Reference number: A60895; PMID:88288254; PMID:2456454  
 A:Accession: A60895  
 A:Molecule type: protein  
 A:Residues: 256-273;292-303,'W',305-306 <SMI>  
 R:Griffith, I.J.; Pollock, J.; Klapper, D.G.; Rogers, B.L.; Nault, A.K.  
 Int. Arch. Allergy Appl. Immunol. 96, 296-304, 1991  
 A>Title: Sequence polymorphism of Amb a I and Amb a II, the major allergens in Ambrosia  
 A:Reference number: A53240; PMID:92234570; PMID:1809687  
 A:Accession: A53240  
 A:Molecule type: mRNA  
 A:Status: preliminary  
 A:Residues: 1-396 <GRI>  
 A:Cross-references: GB:M80558; NID:G166434; PIDN:AAA32665.1; PID:G166435  
 C:Superfamily: pectate lyase LATS9  
 C:Keywords: glycoprotein; pollen  
 F:36/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 64.2%; Score 52; DB 2; Length 396;  
 Best Local Similarity 66.7%; Pred. No. 0.32;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 ||||| :|||  
 Db 113 RPLMIIFKNDMVIRL 127

## RESULT 13

allergen Amb a 1.3 precursor - common ragweed  
 C:Species: Ambrosia artemisiifolia (common ragweed)  
 C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 18-Sep-1998  
 C:Accession: C53240  
 R:Griffith, I.J.; Pollock, J.; Klapper, D.G.; Rogers, B.L.; Nault, A.K.  
 Int. Arch. Allergy Appl. Immunol. 96, 296-304, 1991  
 A>Title: Sequence polymorphism of Amb a I and Amb a II, the major allergens in Ambrosia  
 A:Reference number: A53240; PMID:92234570; PMID:1809687  
 A:Accession: C53240  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-397 <GRI>  
 A:Cross-references: GB:M80560  
 C:Superfamily: pectate lyase LATS9  
 C:Keywords: pollen

Query Match 64.2%; Score 52; DB 2; Length 397;  
 Best Local Similarity 66.7%; Pred. No. 0.32;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 ||||| :|||  
 Db 114 RPLMIIFKNDMVIRL 128

## RESULT 14

allergen Amb a 1.3 - common ragweed  
 C:Species: Ambrosia artemisiifolia (common ragweed)  
 C>Date: 27-Nov-1991 #sequence\_revision 03-Apr-1992 #text\_change 29-Sep-1999  
 C:Accession: C39099  
 R:Reiner, T.; Griffith, I.J.; Kuo, M.; Bond, J.F.; Rogers, B.L.; Klapper, D.G.  
 J. Biol. Chem. 266, 1229-1236, 1991  
 A>Title: Cloning of Amb a I (antigen E), the major allergen family of short ragweed po  
 A:Reference number: A39099; PMID:91093235; PMID:1702434  
 A:Accession: C39099  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-397 <RA>  
 A:Cross-references: GB:M62361; NID:G166440; PIDN:AAA32668.1; PID:G166441  
 C:Superfamily: pectate lyase LATS9  
 C:Keywords: pollen

Query Match 64.2%; Score 52; DB 2; Length 397;  
 Best Local Similarity 66.7%; Pred. No. 0.32;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 ||||| :|||  
 Db 114 RPLMIIFKNDMVIRL 128

## RESULT 15

pectate lyase (EC 4.2.2.2) - trumpet lily  
 S29612  
 C:Species: Lilium longiflorum (trumpet lily)  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 29-Sep-1999  
 C:Accession: S29612  
 R:Kim, S.; Finkel, D.J.; An, G.  
 submitted to the EMBL Data Library, October 1992  
 A>Description: Abundance patterns of lily pollen cDNAs: characterization of three pol  
 A:Reference number: S29611  
 A:Accession: S29612  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-434 <KIM>  
 A:Cross-references: EMBL:Z17328; NID:G19450; PIDN:CAA78976.1; PID:G19451  
 A:Experimental source: cv. Nellie White, mature flower  
 C:Superfamily: pectate lyase LATS9  
 C:Keywords: carbon-oxygen lyase

Query Match 64.2%; Score 52; DB 2; Length 434;  
 Best Local Similarity 66.7%; Pred. No. 0.35;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 ||||| :|||  
 Db 151 RPLMIIFKNDMVIRL 165

Search completed: April 20, 2003, 13:15:38  
 Job time : 7.07895 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds  
(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-28  
Sequence: 1 REPWIFSGNMWIKL 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	81	100.0	374 1	SBP_CRYJA
2	62	76.5	346 1	MPAI_CUPAR
3	62	76.5	367 1	MPAI_JUNAS
4	62	76.5	375 1	MPAI_CHAAB
5	58	71.6	397 1	PEL_TOBAC
6	55	67.9	449 1	PE59_LYCES
7	54	66.7	398 1	MP12_AMBAR
8	53	65.4	392 1	MP14_AMBAR
9	52	64.2	396 1	MP11_AMBAR
10	52	64.2	397 1	MP13_AMBAR
11	52	64.2	434 1	PEL_LILLO
12	48	59.3	397 1	PEL_AMBAR
13	47	58.0	385 1	YEHY_ECOLI
14	46	56.8	398 1	PE56_LYCES
15	46	56.8	404 1	PE61_LYCES
16	41	50.6	915 1	ATCU_VIBCH
17	40	49.4	202 1	KGUA_SCHPO
18	40	49.4	753 1	PEP2_HUMAN
19	39.5	48.8	1043 1	PL1D_MOUSE
20	39	48.1	333 1	YH98_ARCFU
21	39	48.1	372 1	Y412_MYCPN
22	39	48.1	440 1	OMGP_MOUSE
23	39	48.1	440 1	OMGP_HUMAN
24	39	48.1	2504 1	FAS_HUMAN
25	39	48.1	2505 1	FAS_RAT
26	38.5	47.5	1887 1	RPB1_DROME
27	38	46.9	202 1	T2MU_MYCSP
28	38	46.9	227 1	SRPB_SCHPO
29	38	46.9	254 1	PHSC_SALTU
30	38	46.9	475 1	AMT2_ARATH
31	38	46.9	538 1	RKPI_RHIME
32	38	46.9	2511 1	FAS_CHICK
33	37.5	46.3	1852 1	RPB1_CABEL

34	37	45.7	180 1	YPSA_BACSU
35	37	45.7	296 1	YFMW_BACSU
36	37	45.7	351 1	Y630_RICCR
37	37	45.7	386 1	KPBQ_HUMAN
38	37	45.7	387 1	KPBQ_MOUSE
39	37	45.7	387 1	KPBQ_RAT
40	37	45.7	457 1	AROP_ECOLI
41	37	45.7	588 1	CAR7_CANAL
42	37	45.7	656 1	VEXE_SALTU
43	37	45.7	851 1	BF2_CABEL
44	37	45.7	862 1	VG01_HSV1
45	36.5	45.1	267 1	CYSH_PSEAE

## ALIGNMENTS

RESULT 1  
ID SBP\_CRYJA STANDARD; PRT; 374 AA.  
AC P18632;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Sugi basic protein precursor (SBP) (Major allergen Cry j 1) (Cry j I).  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
OK NCBI\_TaxID=3369;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Pollen;  
RX MEDLINE=94183234; PubMed=8135802;  
RA Sone T., Komiya N., Shimizu K., Kusabe T., Morikubo K.,  
RA Kino K.;  
RT "Cloning and sequencing of cDNA coding for Cry j I, a major allergen  
RT of Japanese cedar pollen.";  
RN Biochem. Biophys. Res. Commun. 199;619-625(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pollen;  
RA Namba M., Kurose M., Torioka K., Fukuda S., Kurimoto M.;  
RN Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE OF 22-41.  
RC TISSUE=Pollen;  
RX MEDLINE=89031257; PubMed=3181436;  
RA Tanaka M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inouye S.,  
RA Matsumae T.;  
RT "N-terminal amino acid sequence of a major allergen of Japanese cedar  
RT pollen (Cry j I).";  
RN FEBS Lett. 239;329-332(1988).  
[4]  
RP CARBOHYDRATES.  
RC TISSUE=Pollen;  
RX MEDLINE=95003748; PubMed=7920021;  
RA Hijikata A., Matsumoto I., Kojima K., Ogawa H.;  
RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
RT (Cryptomeria japonica) pollen allergen, Cry j I.";  
RN Int. Arch. Allergy Immunol. 103;198-202(1994).  
[5]  
RP STRUCTURE OF CARBOHYDRATES.  
RC TISSUE=Pollen;  
RX MEDLINE=95332249; PubMed=7608114;  
RA Hino K., Yamamoto S., Sano O., Taniguchi Y., Kohno K., Usui M.,  
RA Fukuda S., Hanzawa H., Hanyama H., Kurimoto M.;  
RT "Carbohydrate structures of the glycoprotein allergen Cry j I from  
RT Japanese cedar (Cryptomeria japonica) pollen.";  
RL J. Biochem. 117;289-295(1995).  
CC -!- PTM: CONTAINS FUCOSE/Xylose-CONTAINING N-LINKED OLIGOSACCHARIDES.  
CC -!- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR  
CC POLLEN. THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
CC -!- MISCELLANEOUS: THE SEQUENCE OF CRY J I FORM A IS SHOWN HERE. FORM

B DIFFERS IN SIX POSITIONS.  
-1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; D26544; BAA05542.1; -;  
 DR EMBL; D26545; BAA05543.1; -;  
 DR EMBL; D34639; BAA07020.1; -;  
 DR PIR; A44773; A44773.  
 DR GlycoSuiteDB; P18632; -;  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Allergen; Glycoprotein; Multigene family; Signal.

FT SIGNAL 1 21  
 FT CHAIN 22 374  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (COMPLEX).  
 FT CARBOHYD 293 293 /FTID=CAR\_000135.  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (COMPLEX).  
 FT CARBOHYD 354 354 /FTID=CAR\_000136.  
 FT VARIANT 12 12 L -> F (IN CRY J 1-B).  
 FT VARIANT 143 143 H -> Y (IN CRY J 1-B).  
 FT VARIANT 202 202 S -> T (IN CRY J 1-B).  
 FT VARIANT 221 221 L -> S (IN CRY J 1-B).  
 FT VARIANT 358 358 Q -> H (IN CRY J 1-B).  
 FT VARIANT 361 361 K -> Q (IN CRY J 1-B).  
 SQ SEQUENCE 374 AA; 40645 MW; 74AB25950248F56F CRC64;

Query Match 100.0%; Score 81; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNNIKL 15  
 DB 87 RPLWIFSGNNNIKL 101

RESULT 2  
 MPAL CUPAR STANDARD; PRT; 346 AA.  
 AC 09SCG9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major pollen allergen Cup a 1.  
 OS Cupressus arizonica.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 CC NCBI\_TaxID=49011;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20571526; PubMed=1122214;  
 RA Acituno B., Del Pozo V., Minguet A., Arrieta I., Cortegano I.,  
 RA Cardaba B., Gallardo S., Rojo M., Palomino P., Lahoz C.;  
 RT "Molecular cloning of major allergen from Cupressus arizonica pollen:  
 RT Cup a 1.";  
 RL Clin. Exp. Allergy 30:1750-1758 (2000).  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; AJ243570; CAB62551.1; -;  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Allergen; Glycoprotein.  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 346 AA; 37589 MW; F1281DCDA1D5DFD0 CRC64;

Query Match 76.5%; Score 62; DB 1; Length 346;  
 Best Local Similarity 80.0%; Pred. No. 0.0021;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNNIKL 15  
 DB 66 KALWIFSGNNNIKL 80

RESULT 3  
 MPAL JUNAS STANDARD; PRT; 367 AA.  
 AC P81254; 09ZNU7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major pollen allergen Jun a 1 precursor.  
 OS Juniperus ashei (Ozark white cedar).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 CC NCBI\_TaxID=13101;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-38; 42-50; 58-80; 88-94;  
 RP 117-124; 134-140; 160-164; 256-263 AND 322-325.  
 RC TISSUE=Pollen;  
 RX MEDLINE=99414163; PubMed=10482835;  
 RA Midoro-Horiuchi T.M., Goldblum R.M., Kurosky A., Wood T.G.,  
 RA Brooks E.G.;  
 RT "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 RT allergen, Jun a 1.";  
 RL J. Allergy Clin. Immunol. 104:613-617 (1999).  
 RN [2]  
 RP SEQUENCE OF 22-50.  
 RC TISSUE=Pollen;  
 RX MEDLINE=99414162; PubMed=10482835;  
 RA Midoro-Horiuchi T., Goldblum R.M., Kurosky A., Goetz D.W.,  
 RA Brooks E.G.;  
 RT "Isolation and characterization of the mountain cedar (Juniperus  
 RT ashei) pollen major allergen, Jun a 1.";  
 RL J. Allergy Clin. Immunol. 104:608-612 (1999).  
 CC -1- DISEASE: THIS PROTEIN CAUSES SEVERAL SEASONAL ALLERGIS RHINITIS  
 CC IN NORTH AMERICA.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; AF106663; AAD03609.1; -;  
 DR EMBL; AF106662; AAD03608.1; -;  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyasef.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Allergen; Glycoprotein; Signal.

FT SIGNAL 1 21  
 FT CHAIN 22 367 MAJOR POLLEN ALLERGEN JUN A 1.  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 367 AA; 39824 MW; FC9B81E675662EF9 CRC64;

Query Match 76.5%; Score 62; DB 1; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.0023;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNMKL 15  
 DB 87 KALWIFSGNNMKL 101

RESULT 4  
 ID MPAL CHAOB STANDARD; PRT; 375 AA.  
 AC 096385;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major pollen allergen Cha o 1 precursor.  
 OS Chamaecyparis obtusa (Japanese cypress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae;  
 OC Chamaecyparis.

NCBI\_TaxID=13415;

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

TISSUE=Pollen;

MEDLINE=96265194; PubMed=8676896;

Suzuki M., Komiyama N., Itoh M., Sone T., Kuno K., Takagi I.,  
 Ohta N.;

"Purification, characterization and molecular cloning of Cha o 1, a  
 major allergen of Chamaecyparis obtusa (Japanese cypress) pollen.";  
 Mol. Immunol. 33:451-460(1996)

-1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D45404; BAA08246.1; -  
 DR InterPro: IPR002022; Amb\_allergen.

DR Pfam; PF00544; pec\_lyase; 1.

DR PRINTS; PR00807; AMBALLERGEN.

KW Allergen; Glycoprotein; Signal.

FT SIGNAL 1 21

FT CHAIN 22 375 MAJOR POLLEN ALLERGEN CHA O 1.

FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 375 AA; 40258 MW; 81CD91DF7066DBF CRC64;

Query Match 76.5%; Score 62; DB 1; Length 375;  
 Best Local Similarity 80.0%; Pred. No. 0.0023;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNMKL 15  
 DB 87 RSLWIFSGNNMKL 101

RESULT 5  
 PEL\_TOBAC

ID PEL\_TOBAC STANDARD; PRT; 397 AA.

AC P40972;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Pectate lyase precursor (EC 4.2.2.2).

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asceridae; easterids; Solanales; Solanaceae; Nicotiana.

NCBI\_TaxID=4097;

SEQUENCE FROM N.A.

STRAIN=cv. Samsun; Tissue=Pollen;

MEDLINE=93043039; PubMed=1421152;

Rogers H.J., Harvey A., Lonsdale D.M.;

"Isolation and characterization of a tobacco gene with homology to  
 pectate lyase which is specifically expressed during  
 microsporogenesis.";

Plant Mol. Biol. 20:493-502 (1992).

-1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 their non-reducing ends.

-1- DEVELOPMENTAL STAGE: MAXIMALLY EXPRESSED LATE IN POLLEN  
 DEVELOPMENT.

-1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X67158; CAA47630.1; -

EMBL; X67159; CAA47631.1; -

EMBL; X61102; CAA43414.1; -

PIR; S26211; S26211.

InterPro: IPR002022; Amb\_allergen.

Pfam; PF00544; pec\_lyase; 1.

PRINTS; PR00807; AMBALLERGEN.

KW Lyase; Signal.

FT SIGNAL 1 25

FT CHAIN 26 397

FT ACT SITE 272

FT CARBOHYD 134 134

FT CARBOHYD 227 227

FT CARBOHYD 156 156

FT CARBOHYD 189 189

FT CARBOHYD 200 200

FT CARBOHYD 202 202

FT CARBOHYD 249 249

SQ SEQUENCE 397 AA; 44351 MW; EF0A82CB5DA7643F CRC64;

Query Match 71.6%; Score 58; DB 1; Length 397;  
 Best Local Similarity 78.6%; Pred. No. 0.012;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLWIFSGNNMKL 15  
 DB 112 PLWIFSGNNMKL 125

Query Match 71.6%; Score 58; DB 1; Length 397;  
 Best Local Similarity 78.6%; Pred. No. 0.012;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLWIFSGNNMKL 15  
 DB 112 PLWIFSGNNMKL 125

RESULT 6  
 ID PEL\_TOBAC STANDARD; PRT; 449 AA.  
 AC P15772;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Probable pectate lyase P59 precursor (EC 4.2.2.2).  
 GN LATS9.

OS Lycopersicon esculentum (Tomato).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids II; Solanales; Solanaceae; Solanum.  
 CC NCBI\_TaxID=4081;  
 CC  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV\_VP36; TISSUE=Anther;  
 RX MEDLINE=91322485; PubMed=1983191;  
 RA Wing R.A., Yamaguchi J., Larabell S.K., Ursin V.M., McCormick S.;  
 RT "Molecular and genetic characterization of two pollen-expressed genes  
 RT that have sequence similarity to pectate lyases of the plant pathogen  
 RT Erwinia";  
 RT Plant Mol. Biol. 14:17-28(1990).  
 CC  
 CC -1- FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND TUBE  
 CC GROWTH.  
 CC  
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X15499; CA35523.1; -;  
 DR PIR; S27098; S27098.  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR Lyase; Multigene family; Signal.  
 KM  
 FT SIGNAL 1 22  
 FT CHAIN 23 449  
 FT ACT SITE 325 325  
 FT CARBOHYD 56 56  
 FT CARBOHYD 80 80  
 FT CARBOHYD 81 81  
 SQ SEQUENCE 449 AA; 50893 MW; 17E3AA13F173B03C CRC64;  
 QY Query Match 67.9%; Score 55; DB 1; Length 449;  
 Best Local Similarity 71.4%; Pred. No. 0.046;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 DB 165 PLWIIIFKGNMIRL 178  
 RESULT 7  
 MP12\_AMBAR STANDARD; PRT; 398 AA.  
 ID MP12\_AMBAR  
 AC P27760;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pollen allergen Amb a 1.2 precursor (Antigen E) (Antigen Amb a I).  
 OS Ambrosia artemisiifolia (Short ragweed).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
 CC Helianthaceae; Ambrosia.  
 CC NCBI\_TaxID=4212;  
 CC  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Pollen;  
 RX MEDLINE=91093235; PubMed=1702434;  
 RA Ratnar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,  
 RA Klapper D.G.;

RT "Cloning of Amb a I (antigen E), the major allergen family of short  
 RT ragweed pollen.";  
 RT J. Biol. Chem. 266:1229-1236(1991).  
 RN  
 RN SEQUENCE FROM N.A., AND VARIANTS.  
 RP TISSUE=Pollen;  
 RX MEDLINE=92234570; PubMed=1809687;  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.;  
 RT "Sequence polymorphism of Amb a I and Amb a II, the major allergens  
 RT in Ambrosia artemisiifolia (short ragweed).";  
 RT Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).  
 CC  
 CC -1- SUBUNIT: MONOMER.  
 CC  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
 CC  
 CC -1- PTM: The N-terminus is blocked.  
 CC  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M62981; AAA32666.1; -;  
 DR EMBL; M80559; AAA32667.1; -;  
 DR PIR; B39099; B39099.  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR Antigen; Allergen; Signal; Multigene family; Polymorphism.  
 KM  
 FT SIGNAL 1 25  
 FT CHAIN 26 398  
 FT VARIANT 345 345  
 FT VARIANT 381 381  
 SQ SEQUENCE 398 AA; 43664 MW; 020DC662D9B7416C CRC64;  
 QY Query Match 66.7%; Score 54; DB 1; Length 398;  
 Best Local Similarity 73.3%; Pred. No. 0.06;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 DB 115 RPLWIIIFKGNMIRL 129  
 RESULT 8  
 MP14\_AMBAR STANDARD; PRT; 392 AA.  
 ID MP14\_AMBAR  
 AC P28744;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pollen allergen Amb a 1.4 precursor (Antigen E) (Antigen Amb a I).  
 OS Ambrosia artemisiifolia (Short ragweed).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
 CC Helianthaceae; Ambrosia.  
 CC NCBI\_TaxID=4212;  
 CC  
 RN SEQUENCE FROM N.A., AND VARIANTS.  
 RP TISSUE=Pollen;  
 RX MEDLINE=92234570; PubMed=1809687;  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.;  
 RT "Sequence polymorphism of Amb a I and Amb a II, the major allergens  
 RT in Ambrosia artemisiifolia (short ragweed).";  
 RT Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).  
 CC  
 CC -1- SUBUNIT: MONOMER.  
 CC  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
 CC  
 CC -1- PTM: The N-terminus is blocked.

```

CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC AMB A I/AMB A II/CRY J I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M80562; AAA32670.1; -.
CC PIR: D53240; D53240.
CC InterPro: IPR002022; Amb allergen.
CC Pfam: PF00544; pec_lyase; 1.
CC PRINTS: PR00807; AMBALLERGEN.
CC Antigen: Allergen; Signal; Multigene family; Polymorphism.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 392 POLLEN ALLERGEN AMB A 1.1.
CC VARIANT 182 188 SHDGPV -> CNDGPA.
CC SEQUENCE 392 AA; 42842 MW; 7B6219C12F365DA9 CRC64;

Query Match
Best Local Similarity 65.4%; Score 53; DB 1; Length 392;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 RPLWIFSGNNMIXL 15
DB 114 RPLWIFARDVIRL 128

RESULT 9
ID MP13_ AMBAR STANDARD; PRT; 396 AA.
AC P27759;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pollen allergen Amb a 1.1 precursor (Antigen E) (Antigen Amb a I).
OS Ambrosia artemisiifolia (short ragweed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
OC Helianthaceae; Ambrosia.
OX NCBI_TaxID=4212;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Pollen;
RX MEDLINE=91093235; PubMed=1702434;
RA Rafnar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,
RA Klapper D.G.;
RT "Cloning of Amb a I (antigen E), the major allergen family of short
RT ragweed pollen."
RL J. Biol. Chem. 266:1229-1236(1991).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC TISSUE=Pollen;
RX MEDLINE=92234570; PubMed=1809687;
RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.;
RT "Sequence polymorphism of Amb a I and Amb a II, the major allergens
RT in Ambrosia artemisiifolia (short ragweed)."
RL Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).
RN [1] SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.
CC -1- PTM: The N-terminus is blocked.
CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC AMB A I/AMB A II/CRY J I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M63116; -; NOT ANNOTATED_CDS.
CC PIR: A39099; A39099.
CC PIR: A53240; A53240.
CC InterPro: IPR002022; Amb allergen.
CC Pfam: PF00544; pec_lyase; 1.
CC PRINTS: PR00807; AMBALLERGEN.
CC Antigen: Allergen; Signal; Multigene family; Polymorphism.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 396 POLLEN ALLERGEN AMB A 1.1.
CC VARIANT 92 92 E -> D.
CC SEQUENCE 396 AA; 42709 MW; 0CE7DDECB2B8841D CRC64;

Query Match
Best Local Similarity 64.2%; Score 52; DB 1; Length 396;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 RPLWIFSGNNMIXL 15
DB 113 RPLWIFERDVIIRL 127

RESULT 10
ID MP13_ AMBAR STANDARD; PRT; 397 AA.
AC P27761;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pollen allergen Amb a 1.3 precursor (Antigen E) (Antigen Amb a I).
OS Ambrosia artemisiifolia (short ragweed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
OC Helianthaceae; Ambrosia.
OX NCBI_TaxID=4212;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen;
RX MEDLINE=91093235; PubMed=1702434;
RA Rafnar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,
RA Klapper D.G.;
RT "Cloning of Amb a I (antigen E), the major allergen family of short
RT ragweed pollen."
RL J. Biol. Chem. 266:1229-1236(1991).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC TISSUE=Pollen;
RX MEDLINE=92234570; PubMed=1809687;
RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.;
RT "Sequence polymorphism of Amb a I and Amb a II, the major allergens
RT in Ambrosia artemisiifolia (short ragweed)."
RL Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).
RN [1] SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.
CC -1- PTM: The N-terminus is blocked.
CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC AMB A I/AMB A II/CRY J I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DR EMBL; M62961; AAA32668.1; -  
 DR EMBL; M80560; AAA32669.1; ALT\_SEQ.  
 DR PIR; C39099; C39099.  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase\_1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Antigen; Allergen; Signal; Multigene family; Polymorphism.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 397 POLLEN ALLERGEN AMB A 1.3.  
 FT VARIANT 48 48 L -> Y.  
 FT SEQUENCE 397 AA; 42928 MW; C8DB4125790DD0A CRC64;  
 SQ  
 Query Match 64.2%; Score 52; DB 1; Length 397;  
 Best Local Similarity 66.7%; Pred. No. 0.13;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 RPLMIFSGNNIKL 15  
 DB 114 RPLMIFKNDVINTL 128

RESULT 11  
 ID - PEL\_LILLO STANDARD; PRT; 434 AA.  
 AC P40973;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Pectate lyase precursor (EC 4.2.2.2).  
 OS Liliaceae; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Liliales; Liliaceae;  
 OC Liliaceae;  
 OC NCBI\_TaxID=4690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nellie white; TISSUE=pollen;  
 RA Kim S.R., Finkel D.J., An G.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Elimination of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; Z17328; CAA78976.1; -  
 CC EMBL; L18911; AAA33398.1; -  
 CC PIR; S29612; S29612.  
 CC InterPro; IPR002022; Amb allergen.  
 CC Pfam; PF00544; pec\_lyase\_1.  
 CC PRINTS; PR00807; AMBALLERGEN.  
 KW Antigen; Signal; Polymorphism.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 397 POLLEN ALLERGEN AMB A 2.  
 FT VARIANT 70 70 N -> D (DETECTED ONLY IN FLOWER DNA).  
 FT VARIANT 138 138 K -> T (DETECTED ONLY IN FLOWER DNA).  
 FT VARIANT 321 321 K -> R (DETECTED ONLY IN FLOWER DNA).  
 FT SEQUENCE 397 AA; 44082 MW; C7861784C9A3DDDD CRC64;  
 SQ  
 Query Match 59.3%; Score 48; DB 1; Length 397;  
 Best Local Similarity 66.7%; Pred. No. 0.66;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 RPLMIFSGNNIKL 15  
 DB 114 RPLMIFKNDVINTL 128

RESULT 13  
 ID - PEL\_LILLO STANDARD; PRT; 397 AA.  
 AC P2762;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pollen allergen Amb a 2 precursor (antigen K) (Antigen Amb a II).  
 OS Ambrosia artemisiifolia (short ragweed).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; easterids II; Asterales; Asteraceae; Asteroideae;  
 OC Helianthaceae; Ambrosia.  
 OC NCBI\_TaxID=4212;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=flower;  
 RC MEDLINE=92013060; PubMed=1717566;  
 RA Rogers B.L., Morgenstern J.P., Griffith I.J., Yu X.-B.,  
 RA Connell C.M., Brauer A.W., King T.P., Garman R.D., Kuo M.-C.C.;  
 RT "Complete sequence of the allergen Amb alpha II. Recombinant  
 RT expression and reactivity with T cells from ragweed allergic  
 RT patients.";  
 RL J. Immunol. 147:2547-2552 (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND VARIANTS.  
 RC TISSUE=pollen;  
 RC MEDLINE=92234570; PubMed=1809687;  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.;  
 RL "Sequence polymorphism of Amb a I and Amb a II, the major allergens  
 RL in Ambrosia artemisiifolia (short ragweed).";  
 RL Int. Arch. Allergy Appl. Immunol. 96:296-304 (1991).  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; M80561; AAA32671.1; -  
 CC PIR; A46469; A46469.  
 CC InterPro; IPR002022; Amb allergen.  
 CC Pfam; PF00544; pec\_lyase\_1.  
 CC PRINTS; PR00807; AMBALLERGEN.  
 KW Antigen; Allergen; Signal; Polymorphism.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 397 POLLEN ALLERGEN AMB A 2.  
 FT VARIANT 70 70 N -> D (DETECTED ONLY IN FLOWER DNA).  
 FT VARIANT 138 138 K -> T (DETECTED ONLY IN FLOWER DNA).  
 FT VARIANT 321 321 K -> R (DETECTED ONLY IN FLOWER DNA).  
 FT SEQUENCE 397 AA; 44082 MW; C7861784C9A3DDDD CRC64;  
 SQ

ID YEHY\_ECOLI STANDARD; PRT; 385 AA.  
 AC P33361; P76435; Rel. 28, Created  
 DT 01-FEB-1994 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Hypothetical ABC transporter permease protein yehy.  
 OS YEHY OR B2130.  
 OC Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / BHD2600;  
 RA Richerich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,  
 RL Church G.M.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.U.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT  
 CC SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE  
 CC SUBSTRATE ACROSS THE MEMBRANE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT  
 CC SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U00007; AAA60493.1; ALT\_INIT.  
 DR EMBL; AE000302; AAC75191.1; -.  
 DR Ecogene; BG12011; yehy.  
 DR InterPro; IPR00515; BPD\_transp.  
 DR Pfam; PF00528; BPD\_transp; 1.  
 DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBER; 1.  
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT TRANSMEM 45 65 POTENTIAL.  
 FT TRANSMEM 70 90 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 225 245 POTENTIAL.  
 FT TRANSMEM 248 268 POTENTIAL.  
 FT TRANSMEM 323 343 POTENTIAL.  
 FT TRANSMEM 347 367 POTENTIAL.  
 FT CONFLICT 346 385 POTENTIAL.  
 SQ SEQUENCE 385 AA; 41138 MW; E57055E3A2B141CC CRC64;  
 -> AAKORH (IN REF. 1).  
 Query Match 58.0%; Score 47; DB 1; Length 385;  
 Best Local Similarity 54.5%; Pred. No. 0.95;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

RESULT 14  
 ID P56\_LYCES STANDARD; PRT; 398 AA.  
 AC P15721;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Probable pectate lyase P56 precursor (EC 4.2.2.2).  
 GN L4156.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 OX NCBI\_Taxid=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. VF36; TISSUE=Anther;  
 RA MEDLINE=91322485; PubMed=1983191;  
 RA Wing R.A., Yamaguchi J., Larebell S.K., Ursin V.M., McCormick S.;  
 RT "Molecular and genetic characterization of two pollen-expressed genes  
 RT that have sequence similarity to pectate lyases of the plant pathogen  
 RT Erwinia.";  
 RL Plant Mol. Biol. 14:17-28(1990).  
 [2]  
 RP REVISIONS.  
 RC STRAIN=cv. VF36; TISSUE=Anther;  
 RA Wing R.A.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND TUBE  
 CC GROWTH.  
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X15500; CA33524.1; -.  
 DR PIR; S08383; S08383.  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Lyase; Multigene family; Signal.  
 FT SIGNAL 1 27 OR 22 (POTENTIAL).  
 FT CHAIN 28 398 PROBABLE PECTATE LYASE P56.  
 FT ACT\_SITE 273 273 POTENTIAL.  
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 398 AA; 44563 MW; 8D676250BD8BC7C8 CRC64;  
 Query Match 56.8%; Score 46; DB 1; Length 398;  
 Best Local Similarity 57.1%; Pred. No. 1.5;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 113 PLWITFARSMRIRL 126  
 QY 2 PLWITFSGNNIKL 15  
 ID 9612\_LYCES STANDARD; PRT; 404 AA.  
 AC P24396;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

```

DE   Style development-specific protein 9612 precursor.
GN   9612.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_Taxid=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. VE36; TISSUE=Pistil;
RX   MEDLINE=91117185; PUBMED=2277637;
RA   Budellier K.A., Smith A.G., Gasser C.S.;
RT   "Regulation of a stylar transmitting tissue-specific gene in
RT   wild-type and transgenic tomato and tobacco.";
RL   Mol. Gen. Genet. 224:183-192(1990).
CC   -!- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPMENT OF THE TRANSMITTING
CC   -!- TISSUE OF THE STYLE AND/OR IN THE EVENTS RELATED TO POLLINATION
CC   SUCH AS SOME ASPECT IN THE FACILITATION OF COMPATIBLE POLLEN TUBE
CC   GROWTH.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN THE PISTIL WHERE IT IS
CC   FOUND IN THE OUTER FIVE LAYERS OF THE STRANDS OF TRANSMITTING
CC   TISSUE WITHIN THE UPPER TWO-THIRDS OF THE STYLE. FOUND AT MUCH
CC   LOWER LEVELS IN THE ANTHERS AND VEGETATIVE ORGANS.
CC   -!- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND DURING ANTHESIS.
CC   -!- SIMILARITY: 54% IDENTICAL TO TOMATO PROTEINS P59 (AC P15722)
CC   AND P56 (AC P15721).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X55193; CAA38979.1; -.
DR   PIR; S12209; S12209.
DR   InterPro; IPR002022; Amb_allergen.
DR   Pfam; PF00544; pec_lyase; 1.
DR   PRINTS; PR00807; AMBALLERGEN.
KW   Signal.
FT   SIGNAL. 1 20
FT   CHAIN 21 404
FT   CARBOHYD 37 37
FT   CARBOHYD 191 191
SQ   SEQUENCE 404 AA; 44298 MW; B26ED69B128D8675 CRC64;
      N-LINKED (GLCNAC. . .) (POTENTIAL).
      N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match 56.8%; Score 46; DB 1; Length 404;
Best Local Similarity 64.3%; Pred. No. 1.5;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 2 PLWIFSGNNIKL 15
DB 120 PLWIFKRDWVIQL 133

```

Search completed: April 20, 2003, 13:07:29  
 Job time : 3.92105 secs



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds

(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524d-28

Perfect score: 81

Sequence: 1 RPLMTIFSGNNMIXL 15

Scoring table: BLOSUM62

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	100.0	374	10	Q8RUR1
2	62	76.5	367	10	Q91LT2
3	62	76.5	367	10	Q91LT1
4	62	76.5	367	10	Q9M4S6
5	62	76.5	367	10	Q9M4S5
6	62	76.5	367	10	Q9M4S4
7	62	76.5	367	10	Q9M4S3
8	62	76.5	367	10	Q9M4S2
9	62	76.5	367	10	Q9M4S1
10	62	76.5	367	10	Q93X16
11	58	71.6	453	10	Q94LR5
12	57	70.4	501	10	Q93Z04
13	57	70.4	542	10	Q9SV40
14	56	69.1	455	10	Q64510
15	56	69.1	463	10	Q9SCP2
16	53	65.4	472	10	Q9LFP5

17	51	63.0	181	10	Q23666	Q23666 arabidopsis
18	51	63.0	392	10	Q9FXD8	Q9FXD8 arabidopsis
19	51	63.0	408	10	Q8W116	Q8W116 arabidopsis
20	51	63.0	438	10	Q43862	Q43862 zea mays (m
21	51	63.0	459	10	Q23665	Q23665 arabidopsis
22	51	63.0	459	10	Q9M9S2	Q9M9S2 arabidopsis
23	50	61.7	274	10	Q24159	Q24159 nicotiana t
24	50	61.7	354	10	Q940Q1	Q940Q1 arabidopsis
25	50	61.7	388	10	Q9M505	Q9M505 vicia vinif
26	50	61.7	409	10	Q93XJ1	Q93XJ1 salix glig
27	50	61.7	431	10	Q23017	Q23017 arabidopsis
28	50	61.7	452	10	Q9LRM5	Q9LRM5 arabidopsis
29	50	61.7	475	10	Q9SRH4	Q9SRH4 arabidopsis
30	49	60.5	226	10	Q23667	Q23667 arabidopsis
31	49	60.5	404	10	Q9SB71	Q9SB71 arabidopsis
32	49	60.5	408	10	Q9CSM8	Q9CSM8 arabidopsis
33	49	60.5	410	10	Q9FMK5	Q9FMK5 arabidopsis
34	49	60.5	432	10	Q93Z25	Q93Z25 arabidopsis
35	49	60.5	501	16	Q9ZC02	Q9ZC02 streptomyce
36	48	59.3	405	10	Q24416	Q24416 fragaria an
37	48	59.3	405	10	Q94PT6	Q94PT6 fragaria an
38	48	59.3	449	10	Q94J42	Q94J42 arabidopsis
39	48	59.3	622	5	Q95SK1	Q95SK1 drosophila
40	47	58.0	324	10	Q8S345	Q8S345 capsicum an
41	47	58.0	385	16	Q8X670	Q8X670 escherichia
42	47	58.0	392	10	Q9FM66	Q9FM66 arabidopsis
43	47	58.0	409	10	Q9LTZ0	Q9LTZ0 arabidopsis
44	46	56.8	401	10	Q24554	Q24554 zinnia eleg
45	46	56.8	407	10	Q9SDW4	Q9SDW4 musa acumin

## ALIGNMENTS

RESULT 1

ID Q8RUR1 PRELIMINARY; PRT; 374 AA.

AC Q8RUR1;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Cry j 1 precursor.

GN Cry j 1.1 OR Cry j 1.2.

OS Cryptomeria japonica (Japanese cedar).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta.

OC Spermatophyta; Coniferales; Cupressaceae; Cryptomeria.

OX NCBI\_TaxID=3369;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE=POLLEN;

RA Futamura N., Shinohara K.;

RT "Isolation and characterization of cDNAs encoding major allergen Cry j

RT 1 from Cryptomeria japonica pollen."

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB081309; BAB86286.1; --

DR EMBL; AB081310; BAB86287.1; --

KW Signal.

FT SIGNAL.

FT CHAIN

SQ SEQUENCE 374 AA; 40720 MW; 90D0085D24BF2BD4 CRC64;

Query Match 100.0%; Score 81; DB 10; Length 374;

Best Local Similarity 100.0%; Pred. No. 5.7e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 RPLMTIFSGNNMIXL 15

Db 87 RPLMTIFSGNNMIXL 101

RESULT 2

ID Q91LT2 PRELIMINARY; PRT; 367 AA.

AC Q9L12; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-2.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21315424; PubMed=11422137;  
 RA Midoro-Horitsu T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Exp. Allergy 31:771-778(2001).  
 DR EMBL; AF151427; AAF80164.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39768 MW; 0A6AC2F1BAF89586 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 : ||||| |||||  
 Db 87 KTLMIIFSGNNIKL 101

## RESULT 3

Q9L11 PRELIMINARY; PRT; 367 AA.  
 ID Q9L11;  
 AC Q9L11;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-1.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21315424; PubMed=11422137;  
 RA Midoro-Horitsu T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Exp. Allergy 31:771-778(2001).  
 DR EMBL; AF151429; AAF80166.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39708 MW; DCBD1981A74E4711 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 : ||||| |||||  
 Db 87 KTLMIIFSGNNIKL 101

## RESULT 4

Q9M4S6 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S6;  
 AC Q9M4S6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.

GN CUPs1.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monnaie R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257491; AAF72625.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39835 MW; B4B3C60108C2C5A3 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 : ||||| |||||  
 Db 87 KALMIIFSGNNIKL 101

## RESULT 5

Q9M4S5 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S5;  
 AC Q9M4S5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monnaie R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257492; AAF72626.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39894 MW; 5D56FC0B3263B741 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLMIIFSGNNIKL 15  
 : ||||| |||||  
 Db 87 KALMIIFSGNNIKL 101

## RESULT 6

Q9M4S4 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S4;  
 AC Q9M4S4;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPs1.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monesalve R.I., Villalba M., Rodriguez R.:  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257493; AAF72627.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39934 MW; 974D3011D74E3D6E CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 RPLMTIFSGNNMTKL 15  
 : ||||| |||||  
 Db 87 KALMTIFSQNNMTKL 101

RESULT 7  
 ID 09M4S3 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S3:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSL.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monesalve R.I., Villalba M., Rodriguez R.:  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39832 MW; B5DFBF5A61C07A53 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 RPLMTIFSGNNMTKL 15  
 : ||||| |||||  
 Db 87 KALMTIFSQNNMTKL 101

RESULT 8  
 ID 09M4S2 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S2:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSL.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monesalve R.I., Villalba M., Rodriguez R.:  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257495; AAF72629.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39819 MW; AE7E055A61C07A53 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 RPLMTIFSGNNMTKL 15  
 : ||||| |||||  
 Db 87 KALMTIFSQNNMTKL 101

RESULT 9  
 ID 093X51 PRELIMINARY; PRT; 367 AA.  
 AC 093X51:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
 DE Putative allergen jun o 1.  
 GN JUN O 1.  
 OS Juniperus oxycedrus (prickly juniper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxID=69008;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=POLLEN;  
 RA TACOVACCI P., Di Felice G., Pini C.:  
 RT "Cloning of Juniperus oxycedrus major allergen."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ293767; CAC48400.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 SQ SEQUENCE 367 AA; 39808 MW; 5D28204DBFD1B907 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 RPLMTIFSGNNMTKL 15  
 : ||||| |||||  
 Db 87 KALMTIFSQNNMTKL 101

RESULT 10  
 ID 093XL6 PRELIMINARY; PRT; 367 AA.  
 AC 093XL6:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Putative allergen Cup a 1 precursor.  
 GN CUP A 1.  
 OS Cupressus arizonica.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=49011;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLLEN;  
 RA Buteroni C., Di Felice G., Pini C.:  
 RT "Cloning of Cupressus arizonica major allergen."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ278498; CAC37790.2; -  
 DR Signal.  
 KW SIGNAL.  
 FT CHAIN 1 21 POTENTIAL.  
 FT SIGNAL 22 367 PUTATIVE ALLERGEN CUP A 1.  
 SQ SEQUENCE 367 AA; 39809 MW; AFE97260423A9F28 CRC64;

Query Match 76.5%; Score 62; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.012; Mismatches 2; Indels 0; Gaps 0;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLMIFSGNNMNL 15  
 :|||||:|||||  
 Db 87 KALMIFSGNNMNL 101

RESULT 11  
 Q94LR5 PRELIMINARY; PRT; 453 AA.  
 ID Q94LR5  
 AC Q94LR5  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Putative peccate lyase.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eubacteriales; Oryzaeae; Oryza.  
 NCBI\_TaxID=4530;  
 RX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Bueli C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,  
 RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Taitlin T.,  
 RA Riggs F., Hsiao J., Zismann V., Blunt S., Pai G., Vanaken S.E.,  
 RA Utterback T.R., Feldblum T.V., Quackenbush J., Salzberg S.L.,  
 RA White O., Frazer C.M.;  
 RL "Oryza sativa chromosome 10 BAC OSUNB0011A08 genomic sequence."  
 RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC034258; AAK54283.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase; 1.  
 DR Lyase.  
 SQ SEQUENCE 453 AA; 48006 MW; 1411BB1A40901DC CRC64;

Query Match 71.6%; Score 58; DB 10; Length 453;  
 Best Local Similarity 64.3%; Pred. No. 0.073;  
 Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 PLMIFSGNNMNL 15  
 :|||||:|||||  
 Db 140 PLMIFSGNNMNL 153

RESULT 12  
 Q93Z04 PRELIMINARY; PRT; 501 AA.  
 ID Q93Z04  
 AC Q93Z04  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE AT3G54920/f28p10\_100 (Putative peccate lyase).  
 GN AT3G54920.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.  
 NCBI\_TaxID=3702;  
 RX [1]  
 RN SEQUENCE FROM N.A.  
 RA Shin P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
 RA Ban J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamuya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Phaf P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis cDNA clones."

Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 (2)  
 RN SEQUENCE FROM N.A.  
 RP Yamada K., Ban J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Tang C.C., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C.C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shin P., Southwick A., Shinozaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene At3G54920 (GI:15233132)."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY058870; AAL24257.1; -  
 DR EMBL: AY074331; AAL67027.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase; 1.  
 DR Lyase.  
 SQ SEQUENCE 501 AA; 53929 MW; 9E0DE36DEF4C7AAB CRC64;

Query Match 70.4%; Score 57; DB 10; Length 501;  
 Best Local Similarity 71.4%; Pred. No. 0.12; Mismatches 2; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PLMIFSGNNMNL 15  
 :|||||:|||||  
 Db 146 PLMIFSGNNMNL 159

RESULT 13  
 Q9SV40 PRELIMINARY; PRT; 542 AA.  
 ID Q9SV40  
 AC Q9SV40  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Peccate lyase-like protein.  
 GN F28P10.100.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.  
 NCBI\_TaxID=3702;  
 RX [1]  
 RN SEQUENCE FROM N.A.  
 RP Choisme N., Robert C., Brothier P., Wincker P., Catolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,  
 RA Lemcke K., Schueller C., Quetier P., Salanoubat M.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL049655; CAB41092.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase; 1.  
 DR PRINTS: PR00807; AMBALERGEN.  
 DR Lyase.  
 SQ SEQUENCE 542 AA; 58573 MW; EB3779D613B72347 CRC64;

Query Match 70.4%; Score 57; DB 10; Length 542;  
 Best Local Similarity 71.4%; Pred. No. 0.13; Mismatches 2; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PLMIFSGNNMNL 15  
 :|||||:|||||  
 Db 146 PLMIFSGNNMNL 159

RESULT 14  
 Q64510 PRELIMINARY; PRT; 455 AA.  
 ID Q64510  
 AC Q64510

DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Putative peccate lyase.  
 GN AT2G02720.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=CV. COLUMBIA;  
 RC MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.J., Barnstead M.E., Feldhym T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Talon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 402:761-768(1999).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=CV. COLUMBIA;  
 RC Lin X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC002521; AAC05350.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW lyase.  
 SQ SEQUENCE 455 AA; 51257 MW; 7C6A35A767A30CA4 CRC64;  
 QY 1 RPLWIFSGNNMIKL 15  
 DB 166 RPLWIFSGNNMIKL 180

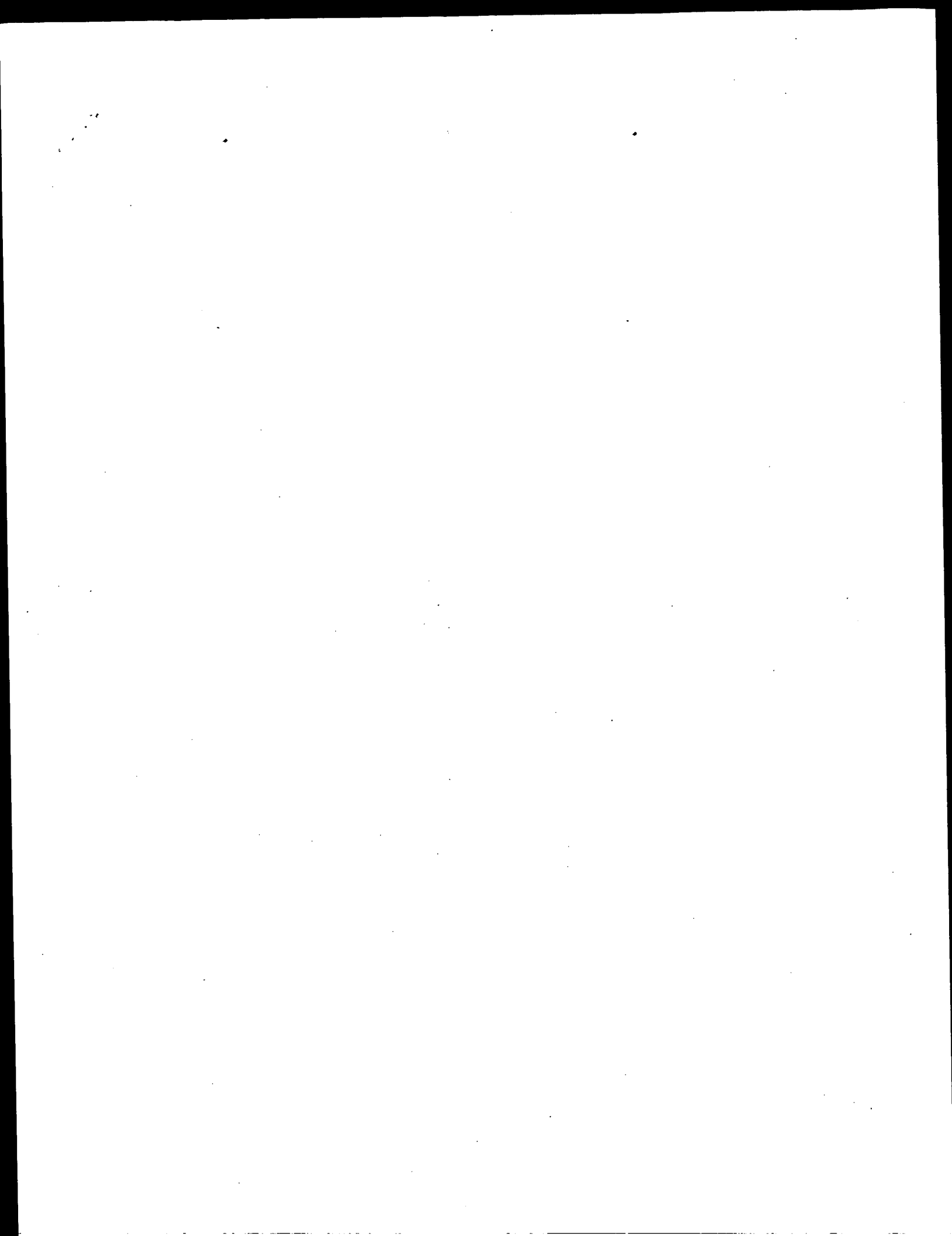
Query Match 69.1%; Score 56; DB 10; Length 455;  
 Best Local Similarity 73.3%; Pred. No. 0.16; 2; Indels 0; Gaps 0;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 15  
 O9SCP2 PRELIMINARY; PRT; 463 AA.  
 ID O9SCP2;  
 AC O9SCP2;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE Peccate lyase-like protein.  
 GN TAD2.120.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP Nyakatura G., Fartmann B., Dauner D., Sterr W., Holland R.,  
 RA Weissegartner M., Mewes H.W., Lemcke K., Mayer K.F.X., Queirer F.,  
 RA Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP EU Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL132958; CAB64222.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.

KW Lyase.  
 SQ SEQUENCE 463 AA; 51225 MW; 3424B0DE4ADBA604 CRC64;  
 QY 2 PLWIFSGNNMIKL 15  
 DB 120 PLWIFSGNNMIKL 133

Query Match 69.1%; Score 56; DB 10; Length 463;  
 Best Local Similarity 71.4%; Pred. No. 0.17;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Search completed: April 20, 2003, 13:12:59  
 Job time : 14.6711 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524D-29  
Perfect score: 79  
Sequence: 1 IFSGNNNIKLKMPMY 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR 73:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	374	2	JC2124
2	79	100.0	374	2	JC2123
3	42	53.2	310	2	T29355
4	40	50.6	245	2	T17617
5	40	50.6	262	2	T43753
6	40	50.6	341	2	H86525
7	40	50.6	341	2	H72097
8	40	50.6	376	1	D89772
9	39	49.4	376	1	C64135
10	39	49.4	765	2	S72278
11	39	49.4	794	2	C86057
12	39	49.4	794	2	A98211
13	38	48.1	283	2	A99294
14	38	48.1	347	2	A99989
15	38	48.1	488	2	C84752
16	38	48.1	757	2	T16609
17	38	48.1	997	2	A60776
18	38	48.1	1344	2	T34188
19	37	46.8	81	2	T44075
20	37	46.8	238	2	E71084
21	37	46.8	275	2	F75102
22	37	46.8	376	2	G84950
23	37	46.8	492	2	AE3539
24	37	46.8	522	2	ISBYSS
25	37	46.8	523	2	B95358
26	37	46.8	542	2	T66728
27	37	46.8	604	2	ACT1083
28	37	46.8	614	2	B86461
29	37	46.8	621	2	F86533

30	37	46.8	621	2	A72091
31	37	46.8	621	2	D81578
32	37	46.8	684	2	T01267
33	37	46.8	768	2	B97725
34	37	46.8	818	2	T08823
35	37	46.8	901	2	F83781
36	37	46.8	928	2	S50578
37	37	46.8	990	2	S23416
38	37	46.8	1225	2	T39255
39	37	46.8	4344	1	A53489
40	37	46.8	4639	1	A54794
41	36	45.6	122	2	A83756
42	36	45.6	161	2	C71044
43	36	45.6	174	2	T22247
44	36	45.6	189	2	T22679
45	36	45.6	233	2	H72648

## ALIGNMENTS

## RESULT 1

JC2124  
major allergen Cry j I precursor (clone PCCI-15) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2124

R/Score, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan

A/Reference number: JC2123; PMID:94183234; PMID:8135802

A/Accession: JC2124

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:D26545; NID:9493633; PIDN:BAA05543.1; PID:9493634

A/Experimental source: pollen

A/Note: the authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase LATS9

C/Keywords: glycoprotein, pollen

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-374/Product: major allergen Cry j I (clone PCCI-15) #status predicted <MAT>

F:158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 79; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. NO. 1e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPMY 15

Db 92 IFSGNNNIKLKMPMY 106

## RESULT 2

JC2123  
major allergen Cry j I precursor (clone PCCI-2-2) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: JC2123; PC2065

R/Score, T.: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan

A/Reference number: JC2123; PMID:94183234; PMID:8135802

A/Accession: JC2123

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:D26544; NID:9493631; PIDN:BAA05542.1; PID:9493632

A/Experimental source: pollen

A/Accession: PC2065

A/Molecule type: protein

A/Residues: 22-53/58-81,219-232;236-258;299-307;346-372 <S02>

A/Note: the authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase LATS9

C/Keywords: glycoprotein, pollen

F:12-21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen Cry I I (clone PCCI-2-2) #status predicted <MAT>  
F:158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 79; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 16-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IFSGNNIKLKMPY 15  
DB 92 IFSGNNIKLKMPY 106

## RESULT 3

T29355  
hypothetical protein R05G6.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2000  
C:Accession: T29355

R:Murphy, J. J. Le, T. T.  
submitted to the EMBL Data Library, May 1996

A:Description: The sequence of C. elegans cosmid R05G6.

A:Reference number: Z20612

A:Accession: T29355

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-310 <MUR>

A:Cross-references: EMBL:U58746; PIDN:AB00622.1; GSPDB:GN00022; CESP:R05G6.7

A:Experimental source: strain Bristol N2; clone R05G6

C:Genetics:

A:Gene: CESP:R05G6.7

A:Map position: 4

A:Introns: 22/1; 69/1; 152/3; 238/3

C:Superfamily: porin

Query Match 53.2%; Score 42; DB 2; Length 310;  
Best Local Similarity 50.0%; Pred. No. 5-9;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNNNNIKLKMPY 15  
DB 57 GNNNNIKLKMPY 68

## RESULT 4

T17617  
hypothetical protein A127R - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C:Accession: T17617

R:Graves, M. V.; Van Etten, J. L.  
submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: T17617

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-245 <GRA>

A:Cross-references: EMBL:U42580; NID:G4028896; PIDN:AAC36435.1

A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:

A:Note: A127R

C:Superfamily: Chlorella virus PBCV-1 hypothetical protein A127R

Query Match 50.6%; Score 40; DB 2; Length 245;  
Best Local Similarity 72.7%; Pred. No. 11;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 FSGNNIKLKMPY 12  
DB 62 FSGNNIKLKMPY 72

## RESULT 5

T43753  
probable COI intron protein [similarity] - alme mold (Dictyostelium discoideum) mito-  
C:Species: Dictyostelium discoideum  
C:Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 06-Oct-2000  
C:Accession: T43753  
R:Ogawa, S.; Yoshino, R.; Angata, K.; Pi, M.; Yamoto, M.; Kuroe, K.; Matsuo, K.; Mori  
submitted to the EMBL Data Library, December 1996  
A:Description: The mitochondrial DNA of Dictyostelium discoideum. Complete sequence,  
A:Reference number: Z22666  
A:Accession: T43753  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-262 <OGA>

A:Cross-references: EMBL:AB000109; PIDN:BA078057.1

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: COI intron 9 protein; COI intron 9 protein homology

F:12-241/Domain: COI intron 9 protein homology <CI9>

Query Match 50.6%; Score 40; DB 2; Length 262;  
Best Local Similarity 61.5%; Pred. No. 12;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 SGNNIKLKMPY 15  
DB 31 SGNNIKLKMPY 43

## RESULT 6

H86525  
dipeptide transporter Arpase [imported] - Chlamydomonas reinhardtii (strain J138)

C:Species: Chlamydomonas reinhardtii

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: H86525

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydomonas reinhardtii J138.

A:Reference number: AB6491; PMID:20330349; PMID:10871362

A:Accession: H86525

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-341 <STO>

A:Cross-references: GB:BA000008; NID:G9978654; PIDN:BA098490.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: dppf\_1

C:Superfamily: probable ABC-type transport protein abc; ATP-binding cassette homology

Query Match 50.6%; Score 40; DB 2; Length 341;  
Best Local Similarity 57.1%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 IFSGNNIKLKMPY 14  
DB 296 IFSGNNIKLKMPY 309

## RESULT 7

B72097  
dipeptide transporter Arpase - Chlamydomonas reinhardtii (strain CML029)

C:Species: Chlamydomonas reinhardtii

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C:Accession: B72097

R:Kalan, S.; Mitchell, W.; Marathe, R.; Lamm, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.

A:Reference number: A72000; PMID:99206606; PMID:10192388

A:Accession: B72097

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-341 <ARN>

A:Cross-references: GB:AE001613; GB:AE001363; NID:G4376550; PIDN:AD018429.1; PID:G43



A:Experimental source: strain CML029

C:Genetics:

A:Gene: dppf\_1

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match

Best Local Similarity 50.6%; Score 40; DB 2; Length 341;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IFSGNNNKKMKMPY 14

Db 296 ILSGNINLFRKSPM 309

RESULT 8

hypothetical protein SA016 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: D89772

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; PMID:21311952; PMID:11418146

A:Accession: D89772

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-578 <KDR>

A:Cross-references: GB:BA000018; PID:g13700036; PIDN:BA81335.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA016

Query Match

Best Local Similarity 50.6%; Score 40; DB 2; Length 578;

Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IFSGNNNKKMKMPY 15

Db 264 VRSKALNITKLPFH 278

RESULT 9

C64135

ribonucleoside-diphosphate reductase (EC 1.17.4.1) beta chain - Haemophilus influenzae

C:Species: Haemophilus influenzae

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: C64135

R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weisman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

A:Authors: Grehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.

; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; PMID:95350630; PMID:7542800

A:Accession: C64135

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-376 <TRG>

A:Cross-references: GB:032839; GB:I42023; NID:g1574508; PIDN:AC23306.1; PID:g1574510; T

C:Superfamily: herpesvirus ribonucleoside-diphosphate reductase small chain

C:Keywords: DNA replication; iron; metalloprotein; oxidoreductase

F:85,116,119,205,239,242/Binding site: 2Fe-O cluster (Asp, Glu, His, Glu, Glu, His) #sta

Query Match

Best Local Similarity 49.4%; Score 39; DB 1; Length 376;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 FSGNNNKKMKMPY 15

Db 6 FSGNNNKKMKMPY 19

RESULT 10

S72278

ATP-dependent Clp proteinase (EC 3.4.21.-) homolog - Plasmodium falciparum

C:Species: Plasmodium falciparum

C>Date: 14-Apr-1998 #sequence\_revision 24-Apr-1998 #text\_change 20-Jun-2000

C:Accession: S72278; S78483

R:Wilson, R.J.M.; Denny, P.W.; Pretser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; W

J. Mol. Biol. 261, 155-172, 1996

A:Title: Complete gene map of the plactid-like DNA of the malaria parasite Plasmodium

A:Reference number: S72277; PMID:96346169; PMID:8757284

A:Accession: S72278

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-765 <WIL>

A:Cross-references: EMBL:X95276

R:Wilson, R.J.M.

submitted to the EMBL Data Library, January 1996

A:Reference number: S78483

A:Accession: S78483

A:Molecule type: DNA

A:Residues: 1-95, 'N', 96-765 <WIL>

A:Cross-references: EMBL:X95276; NID:g1171591; PIDN:CAA64596.1; PID:g1171612

C:Genetics:

A:Gene: clpC

A:Note: this apparently degenerate plactid is referred to as the apicoplast

C:Superfamily: endopeptidase Clp ATP-binding chain

C:Keywords: hydrolase; plactid; serine proteinase

Query Match

Best Local Similarity 49.4%; Score 39; DB 2; Length 765;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IFSGNNNKKMKMPY 11

Db 143 IFSGNNNKKMKMPY 153

RESULT 11

C86057

hypothetical protein Z5214 [imported] - Escherichia coli (strain O157:H7, substrain EF

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: C86057

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh

iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; DiMaio, E.; Potamousis, K.; Apodac

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Accession: C86057

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-794 <STO>

A:Cross-references: GB:AE001174; NID:g12518567; PIDN:AGS8919.1; GSPDB:GN00145; UMG

A:Experimental source: strain O157:H7, substrain EDJ933

C:Genetics:

A:Gene: Z5214

Query Match

Best Local Similarity 49.4%; Score 39; DB 2; Length 794;

Matches 8; Conservative 5; Mismatches 2; Indels 2; Gaps 1;

Qy 1 IFSGNNNKKMKMPY 15

Db 456 ILSGNLANTQVKIRPV 472

RESULT 12

A98211

hypothetical protein Eca657 [imported] - Escherichia coli (strain O157:H7, substrain R

C:Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C/Accession: A98211  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Ganaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene  
 A/Reference number: A99629; PMID:21156231; PMID:11258796  
 A/Accession: A98211  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-794 <HAY>  
 A/Cross-references: GB:BA000007; PIDN:BAH38080.1; PID:g13364132; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RMD 0509952  
 C/Genetics:  
 A/Gene: EC84657

Query Match 49.4%; Score 39; DB 2; Length 794;  
 Best Local Similarity 47.1%; Pred. No. 61;  
 Matches 8; Conservative 5; Mismatches 2; Indels 2; Gaps 1;

OY 1 IFSGNM--NITKMPY 15  
 ||||| : : : : :  
 Db 456 ILSGNLANTQVKIPY 472

## RESULT 13

A99294

hypothetical protein SS01374 [imported] - *Sulfolobus solfataricus*  
 C/Species: *Sulfolobus solfataricus*  
 C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C/Accession: A99294  
 R/She, Q.; Singh, R.K.; Cafaloni, F.; Zivanovic, Y.; Allard, G.; Anayez, M.J.; Chan-  
 Jiong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.  
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A/Description: *Sulfolobus solfataricus* complete genome.  
 A/Reference number: A99139  
 A/Accession: A99294  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-283 <KUR>  
 A/Cross-references: GB:AE006641; NID:g13814586; PIDN:AAK41608.1; GSPDB:GN00155  
 C/Genetics:  
 A/Gene: SS01374

Query Match 48.1%; Score 38; DB 2; Length 283;  
 Best Local Similarity 38.5%; Pred. No. 30;  
 Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 3 SGNMNIKMKPMY 15  
 : : : : :  
 Db 159 AGNLSIPKIPILH 171

## RESULT 14

A99989

Cyelin B [imported] - *Gullardia theta* nucleomorph  
 C/Species: nucleomorph *Gullardia theta*

A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
 C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
 C/Accession: A99989  
 R/Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei-  
 Nature 410, 1091-1096, 2001  
 A/Title: The highly reduced genome of an enslaved algal nucleus.  
 A/Reference number: A99082; PMID:11323671; PMID:11323671  
 A/Accession: A99989  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-347 <DOU>  
 A/Cross-references: GB:AF15818; NID:g13794469; PIDN:AAK39844.1; GSPDB:GN00150  
 C/Genetics:  
 A/Gene: cycB  
 A/Map position: 1

A/Genome: nucleomorph  
 C/Keywords: nucleomorph

Query Match 48.1%; Score 38; DB 2; Length 347;  
 Best Local Similarity 42.9%; Pred. No. 37;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 2 FSGNMNIKMKPMY 15  
 ||||| : : : : :  
 Db 279 FSGNISIKITIMIF 292

## RESULT 15

C84752

hypothetical protein At2g34090 [imported] - *Arabidopsis thaliana*  
 C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C/Accession: C84752  
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A/Reference number: A84420; PMID:20083487; PMID:10617197  
 A/Accession: C84752  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-488 <STO>  
 A/Cross-references: GB:AE002093; NID:g2342730; PIDN:AAE67628.1; GSPDB:GN00139  
 C/Genetics:  
 A/Gene: At2g34090  
 A/Map position: 2  
 C/Superfamily: Saccharomyces cerevisiae probable membrane protein YNL011c

Query Match 48.1%; Score 38; DB 2; Length 488;  
 Best Local Similarity 58.3%; Pred. No. 54;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 GNMNIKMKPMY 15  
 ||||| : : : : :  
 Db 380 GDTNIRLKNPRTY 391

Search completed: April 20, 2003, 13:15:40  
 Job time : 8.07895 secs

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-29

Sequence: 1 IFSGNNMIKMPMT 15

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	374	1 SBP_CRYJA	P18632 cryptomeria
2	68	86.1	367	1 MPAL_JUNAS	P81294 juniperus a
3	64	81.0	346	1 MPAL_CUPAR	O96c99 cupressus a
4	60	75.9	375	1 MPAL_CABEL	O96385 chamaecypar
5	42	53.2	283	1 PORI_CABEL	Q21752 caenohabdi
6	39	49.4	375	1 RIR2_HABIN	P43755 haemophilus
7	38	48.1	111	1 FER_BUCAP	O51882 buchnera ap
8	37	46.8	376	1 RIR2_BUCAI	P57275 buchnera ap
9	37	46.8	522	1 PDI_YEAST	P17967 saccharomyc
10	37	46.8	621	1 Y344_CHLBN	O9k275 chlamydia p
11	37	46.8	928	1 PTP3_YEAST	P40048 saccharomyc
12	37	46.8	986	1 EP1B_STREP	P20195 staphylococ
13	37	46.8	4344	1 DYHC_EMENT	P45444 emericella
14	37	46.8	4639	1 DYHC_DROME	P37276 drosophila
15	36	45.6	268	1 TRJ3_YEAST	O99394 saccharomyc
16	36	45.6	424	1 TRPB_THRAC	O9hkd2 thermoplas
17	36	45.6	505	1 YML8_YEAST	Q03210 saccharomyc
18	36	45.6	770	1 NASE_BACSU	P42433 bacillus su
19	36	45.6	837	1 HFC3_HABIN	P45998 haemophilus
20	36	45.6	946	1 YIA2_YEAST	P40559 saccharomyc
21	36	45.6	1080	1 NRK1_YEAST	P28692 saccharomyc
22	36	45.6	1286	1 YKVS_YEAST	P28273 saccharomyc
23	36	45.6	1536	1 Y984_THEMA	O9x079 thermotoga
24	36	45.6	4349	1 DYHC_FUSSO	P78716 fusarium so
25	35	44.3	327	1 SRP8_SCHPO	O13950 schizosach
26	35	44.3	367	1 YC36_HABIN	P44132 haemophilus
27	35	44.3	492	1 SYNM_YEAST	P25345 saccharomyc
28	35	44.3	498	1 POLG_DEN19	P27909 dengue viru
29	35	44.3	555	1 POLG_DEN17	P29983 dengue viru
30	35	44.3	557	1 SYD_MYCPN	P75068 mycoplasma
31	35	44.3	581	1 FIFB_ADEOS	P11818 human adeno
32	35	44.3	791	1 POLG_DEN1C	P27913 dengue viru
33	35	44.3	792	1 POLG_DEN1B	P27910 dengue viru

## ALIGNMENTS

RESULT 1	ID	SBP_CRYJA	STANDARD;	PRT;	374 AA.
AC	P18632;				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Sugi basic protein precursor (SBP) (Major allergen Cry j 1) (Cry j I).				
OS	Cryptomeria japonica (Japanese cedar)				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.				
OK	NCBI_Taxid=3369;				
RN	(1)				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Pollen;				
RA	MEDLINE=94183234; PubMed=8135802;				
RX	Sone T., Komiya N., Shimizu K., Kusakabe T., Morikubo K.,				
RA	Kino K.;				
RT	"Cloning and sequencing of cDNA coding for Cry j I, a major allergen				
RT	of Japanese cedar pollen."				
RL	Biochem. Biophys. Res. Commun. 199;619-625 (1994).				
RN	(2)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pollen;				
RA	Namua M., Kurose M., Torioge K., Fukuda S., Kurimoto M.;				
RL	Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.				
RN	(3)				
RP	SEQUENCE OF 22-41.				
RC	TISSUE=Pollen;				
RX	MEDLINE=89031257; PubMed=3181436;				
RA	Tanai M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inoue S.,				
RT	Matuhashi T.;				
RT	"N-terminal amino acid sequence of a major allergen of Japanese cedar				
RT	pollen (Cry j I)."				
RL	FEBS Lett. 239;329-332 (1988).				
RN	(4)				
RP	CARBOHYDRATES.				
RC	TISSUE=Pollen;				
RX	MEDLINE=95003748; PubMed=7920021;				
RA	Hijikata A., Matsumoto I., Kojima K., Ogawa H.;				
RT	"Antigenicity of the oligosaccharide moiety of the Japanese cedar				
RT	(Cryptomeria japonica) pollen allergen, Cry j I."				
RL	Int. Arch. Allergy Immunol. 105;198-202 (1994).				
RN	(5)				
RP	STRUCTURE OF CARBOHYDRATES.				
RC	TISSUE=Pollen;				
RX	MEDLINE=95332249; PubMed=7608114;				
RA	Hino K., Yamamoto S., Sano O., Taniguchi Y., Kohno K., Usui M.,				
RT	Fukuda S., Hanzawa H., Hanayama H., Kurimoto M.;				
RT	"Carbohydrate structures of the glycoprotein allergen Cry j I from				
RT	Japanese cedar (Cryptomeria japonica) pollen."				
RL	J. Biochem. 117;289-295 (1995).				
CC	-1- PPM: CONTAINS FUCOSE/YULOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.				
CC	-1- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR				
CC	POLLEN, THE MOST COMMON POLLEN ALLERGEN IN JAPAN.				
CC	-1- MISCELLANEOUS: THE SEQUENCE OF CRY J I FORM A IS SHOWN HERE. FORM				

P27912 dengue viru  
P36418 dictyosteli  
P17763 dengue viru  
O61026 mus masculu  
O15596 homo sapien  
P14046 rattus norv  
O9440 homo sapien  
O99191 sus scrofa  
P33478 d genome po  
P09091 saccharomyc  
Q91917 arabidopsi

B DIFFERS IN SIX POSITIONS.  
 -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; D26544; BAA0542.1; -  
 CC EMBL; D26545; BAA0543.1; -  
 CC EMBL; D34639; BAA07020.1; -  
 CC PIR; A44773; A44773.  
 CC GlycoSuiteDB; P18632; -  
 CC InterPro; IPR02022; Amb\_allergen.  
 CC Pfam; PF00544; pec\_lyase; 1.  
 CC PRINTS; PR00807; AMBALLERGEN.  
 CC Allergen; Glycoprotein; Multigene family; signal.

-----  
 KW SIGNAL 1 21  
 FT CHAIN 22 374 SUGI BASIC PROTEIN. (POTENTIAL).  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (COMPLEX).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (COMPLEX).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (COMPLEX).  
 FT VARIANT 12 12 L -> F (IN CRY J 1-B).  
 FT VARIANT 143 143 H -> Y (IN CRY J 1-B).  
 FT VARIANT 202 202 S -> T (IN CRY J 1-B).  
 FT VARIANT 221 221 L -> S (IN CRY J 1-B).  
 FT VARIANT 358 358 Q -> H (IN CRY J 1-B).  
 FT VARIANT 361 361 K -> Q (IN CRY J 1-B).  
 SQ SEQUENCE 374 AA; 40645 MW; 74AB25950248F56F CRC64;

Query Match 100.0%; Score 79; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNITKIKMPY 15  
 DB 92 IFSGNNITKIKMPY 106

-----  
 AC P81294; Q922U7; STANDARD; PRT; 367 AA.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major pollen allergen Jun a 1 precursor.  
 OS Juniperus ashei (Ozark white cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_Taxid=13101;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-38; 42-50; 58-80; 88-94;  
 RP 117-124; 134-140; 160-164; 256-263 AND 322-325.  
 RC TISSUE=pollen;  
 RA MEDLINE=99414163; PubMed=10482836;  
 RA MIDRO-Horvut T.M., Goldblum R.M., Kurosky A., Wood T.G.,  
 RA Brooks E.G.;  
 RT "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 RT allergen, Jun a 1,";  
 RT J Allergy Clin. Immunol. 104:613-617(1999).  
 RL [2]  
 RN SEQUENCE OF 22-50.  
 RP TISSUE=pollen;  
 RC MEDLINE=99414162; PubMed=10482835;  
 RA MIDRO-Horvut T., Goldblum R.M., Kurosky A., Goetz D.W.,

RA Brooks E.G.;  
 RT "Isolation and characterization of the mountain cedar (Juniperus  
 RT ashei) pollen major allergen, Jun a 1,";  
 RL J. Allergy Clin. Immunol. 104:608-612(1999).  
 CC -1- DISEASE: THIS PROTEIN CAUSES SEVERAL SEASONAL ALLERGIC RHINITIS  
 CC IN NORTH AMERICA.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; AF106663; AAD03609.1; -  
 CC EMBL; AF106662; AAD03608.1; -  
 CC InterPro; IPR02022; Amb\_allergen.  
 CC Pfam; PF00544; pec\_lyase; 1.  
 CC PRINTS; PR00807; AMBALLERGEN.  
 CC Allergen; Glycoprotein; signal.

-----  
 KW SIGNAL 1 21  
 FT CHAIN 22 367 MAJOR POLLEN ALLERGEN JUN A 1.  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 367 AA; 39824 MW; FC9B81B67562E49 CRC64;

Query Match 86.1%; Score 68; DB 1; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 1.3e-05;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IFSGNNITKIKMPY 15  
 DB 92 IFSGNNITKIKMPY 106

-----  
 AC P81294; Q922U7; STANDARD; PRT; 346 AA.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major pollen allergen Cup a 1.  
 OS Cupressus arizonica.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_Taxid=49011;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=20571526; PubMed=11122214;  
 RA Acetunio E., Del Pozo V., Minguez A., Arrieta I., Cortegano I.,  
 RA Cardaba B., Gallardo S., Rojo M., Palomino P., Lahoz C.;  
 RT "Molecular cloning of major allergen from Cupressus arizonica pollen:  
 RT Cup a 1,";  
 RT Clin. Exp. Allergy 30:1750-1758(2000).  
 RL -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; AJ243570; CAB62551.1; -  
 CC InterPro; IPR02022; Amb\_allergen.  
 CC Pfam; PF00544; pec\_lyase; 1.  
 CC PRINTS; PR00807; AMBALLERGEN.

KM Allergen; Glycoprotein.  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 346 AA; 37589 MW; F1281DDDA1D5DD0 CRC64;

Query Match 81.0%; Score 64; DB 1; Length 346;  
 Best Local Similarity 80.0%; Pred. No. 7.3e-05;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IFSGNMNIKLKMPY 15  
 DB 71 IFSGNMNIKLKMPY 85

RESULT 4  
 MPAL CHAQB STANDARD; PRT; 375 AA.

ID MPAL CHAQB STANDARD; PRT; 375 AA.  
 AC 096385;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major pollen allergen Cha o 1 precursor.  
 OS Chamaecyparis obtusa (Japanese cypress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae;  
 OC Chamaecyparidaceae.  
 NC NCBI\_Taxid=13415;

RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC TISSUE=Pollin;  
 RX MEDLINE=96265194; PubMed=4676896;  
 RA Suzuki M., Komiyama N., Itoh H., Sone T., Kuno K., Takagi I.,  
 RT "Purification, characterization and molecular cloning of Cha o 1, a  
 major allergen of Chamaecyparis obtusa (Japanese cypress) pollen."  
 RL Mol. Immunol. 33:451-460(1996).  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; IMA5404; BAA08246.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase.1  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Allergen; Glycoprotein; Signal.

FT SIGNAL 1 21 MAJOR POLLEN ALLERGEN CHA O 1.  
 FT CHAIN 22 375  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 375 AA; 40258 MW; 81CD91DF7066DBF CRC64;

Query Match 75.9%; Score 60; DB 1; Length 375;  
 Best Local Similarity 73.3%; Pred. No. 0.00040;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IFSGNMNIKLKMPY 15  
 DB 92 IFSGNMNIKLKMPY 106

RESULT 5  
 PORT CAEBL

ID PORT CAEBL STANDARD; PRT; 283 AA.  
 AC 021752;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable voltage-dependent anion-selective channel.  
 GN R05G6.7.  
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NC NCBI\_Taxid=6239;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2.  
 RA Murray J., Je T.T.  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.

RA Waterston R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE CELL MEMBRANE THAT ALLOWS  
 CC DIFFUSION OF SMALL HYDROPHILIC MOLECULES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE (BY  
 CC SIMILARITY).  
 CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIZED BETA-SHEETS.  
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U58746; AAB00622.2; -  
 DR WormPep; R05G6.7; CE29443.  
 DR InterPro; IPR01925; Porin Euk.  
 DR Pfam; PF01459; Euk porin; 1.  
 DR PRINTS; PR00185; EUKARYOTIC.  
 DR PROSITE; PS00558; EUKARYOTIC\_PORIN; FALSE\_NEG.  
 KW Outer membrane; Porin; Mitochondrion.  
 SQ SEQUENCE 283 AA; 29960 MW; 060141EBA5060290 CRC64;

Query Match 53.2%; Score 42; DB 1; Length 283;  
 Best Local Similarity 50.0%; Pred. No. 1.1;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNMNITKLKMPY 15  
 DB 57 GNMNITKLKMPY 68

RESULT 6  
 RIR2 HAEIN STANDARD; PRT; 375 AA.  
 AC P43755;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribonucleoside-diphosphate reductase beta chain (EC 1.17.4.1)  
 DE (Ribonucleoside reductase small subunit).  
 GN NRDB OR H11660.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.

CC NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KM20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

```

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geisbrenner N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512 (1995).
CC -1- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY
CC FOR DNA SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC thioedoxin + H(2)O = ribonucleoside diphosphate + reduced
CC thioedoxin.
CC -1- COFACTOR: BINDS 2 IRON IONS (BY SIMILARITY).
CC -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U32839; AAC23306.1; -.
CC DR HSSP; P00453; IXIK.
CC DR TIGR; H11600; -.
CC DR InterPro; IPR000358; Ribonuc_redctase.
CC DR Pfam; PF00268; ribonuc_red_sm; 1.
CC DR PROSITE; PS00368; RIBORED_SMALL; 1.
CC KW Oxidoreductase; DNA replication; Iron; Complete proteome.
CC FT METAL 84 84 IRON 1 (BY SIMILARITY).
CC FT METAL 115 115 IRON 1 AND 2 (BY SIMILARITY).
CC FT METAL 118 118 IRON 1 (BY SIMILARITY).
CC FT METAL 204 204 IRON 2 (BY SIMILARITY).
CC FT METAL 238 238 IRON 2 (BY SIMILARITY).
CC FT METAL 241 241 IRON 2 (BY SIMILARITY).
CC FT ACT SITE 122 122 BY SIMILARITY.
CC SQ SEQUENCE 375 AA; 43194 MW; 7BA30D8C96A5BB35 CRC64;
CC -----
Query Match 49.4%; Score 39; DB 1; Length 375;
Best Local Similarity 57.1%; Pred. No. 6.1;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

RT "Sequence analysis of a 34.7-kb DNA segment from the genome of
RT Buchnera aphidicola (endosymbiont of aphids) containing groEL, dnaA,
RT the acp operon, gldA, and rho."
RL Curr. Microbiol. 36:158-163 (1998).
CC -1- FUNCTION: FERREDOXIN ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -1- COFACTOR: BINDS 1 2FE-2S CLUSTER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ADRENODOXIN / PUTIDREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF008210; AAC38120.1; -.
CC DR HSSP; P00257; 1E6E.
CC DR InterPro; IPR001055; Adrenodoxin.
CC DR Pfam; PF00111; fer2; 1.
CC DR PRINTS; PR00355; ADRENODOXIN.
CC DR PROSITE; PS00814; ADX; 1.
CC KW Electron transport; Iron-sulfur.
CC FT METAL 42 42 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC FT METAL 48 48 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC FT METAL 51 51 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC FT METAL 87 87 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC SQ SEQUENCE 111 AA; 12458 MW; A0DB3A1209A935D CRC64;
CC -----
Query Match 48.1%; Score 38; DB 1; Length 111;
Best Local Similarity 41.7%; Pred. No. 2.3;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 FSGNNMNIKLMKMPY 15
DB 5 FSGNNMNIKLMKMPY 18

```

```

OY 4 GNNMNIKLMKMPY 15
DB 92 GNNMNIKLMKMPY 103

```

CC SMALL CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; AP001118; BAB12895.1; -

DR HSSP; P00453; IXTK.

DR InterPro; IPR000358; Ribonuc[redacted] redctse.

DR Pfam; PF00268; ribonuc\_red\_sm; 1.

DR PROSITE; PS00368; RIBORED\_SMALL; 1.

KW Oxidoreductase; DNA replication; Iron; Complete proteome.

FT METAL 85 85 IRON 1 (BY SIMILARITY).

FT METAL 116 116 IRON 1 AND 2 (BY SIMILARITY).

FT METAL 119 119 IRON 1 (BY SIMILARITY).

FT METAL 205 205 IRON 2 (BY SIMILARITY).

FT METAL 239 239 IRON 2 (BY SIMILARITY).

FT METAL 242 242 IRON 2 (BY SIMILARITY).

FT ACT\_SITE 123 123 BY SIMILARITY.

FT SEQUENCE 376 AA; 44145 MW; BECC1B02F77CBB32 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 376;

Best Local Similarity 53.3%; Pred. No. 15;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IFSGNMNKLKMPY 15

Db 5 IFSKKNNQLKEPMF 19

RESULT 9

PDI\_YEAST

ID PDI\_YEAST STANDARD; PRT; 522 AA.

AC P17967;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Thioredoxin-

DE related glycoprotein 1).

GN PDI1 OR MFP1 OR TRG1 OR YCL043C OR YCL43C OR YCL313.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=X2180-1A, and TM5;

RX MEDLINE=92105067; PubMed=1761527;

RA Tachikawa H., Miura T., Katakura Y., Mizunaga T.;

RT "Molecular structure of a yeast gene, PDI1, encoding protein

RT disulfide isomerase that is essential for cell growth.";

RL J. Biochem. 110:306-313(1991).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=91289690; PubMed=2063627;

RA Scherens B., Dubois E., Messenguy F.;

RT "Determination of the sequence of the yeast YCL313 gene localized on

RT chromosome III. Homology with the protein disulfide isomerase (PDI

RT gene product) of other organisms.";

RL Yeast 7:185-193(1991).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=91239586; PubMed=1840656;

RA Lamantia M., Miura T., Tachikawa H., Kaplan H.A., Lennarz W.J.,

RA Mizunaga T.;

RT "Glycosylation site binding protein and protein disulfide isomerase

RT are identical and essential for cell viability in yeast.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:4453-4457(1991).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=92105117; PubMed=1761554;

RA Gauthier R., Brauer C., Janetzky B., Foerster H.H., Ehbrecht I.M.,

RA Lehle L., Kuentzel H.;

RT "The Saccharomyces cerevisiae TRG1 gene is essential for growth and

RT encodes a luminal endoplasmic reticulum glycoprotein involved in the

RT maturation of vacuolar carboxypeptidase.";

RL U. Biol. Chem. 266:24557-24563(1991).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=92397595; PubMed=1523890;

RA Scherens B., Messenguy F., Giot D., Dubois E.;

RT "The complete sequence of a 9,543 bp segment on the left arm of

RT chromosome III reveals five open reading frames including glucokinase

RT and the protein disulfide isomerase.";

RL Yeast 8:577-586(1992).

RN [6]

RP SEQUENCE FROM N.A.

RX MEDLINE=92104510; PubMed=1761235;

RA Farquhar R., Honey N., Murant S.J., Boesier P., Schultz L.,

RA Montgomerie D., Ellis R.W., Freedman R.B., Tuile M.F.;

RT "Protein disulfide isomerase is essential for viability in

RT Saccharomyces cerevisiae.";

RL Gene 108:81-89(1991).

CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING

CC DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL

CC HYDROXYMATION AND TRIGLYCERIDE TRANSFER.

CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and

CC interchain disulfide bonds in proteins to form the native

CC structures.

CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (Potential).

CC -1- PTM: The N-terminus is blocked.

CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X52313; CAA36550.1; -

DR EMBL; X57712; CAA40883.1; -

DR EMBL; M62815; AAA34948.1; -

DR EMBL; M76982; AAA35169.1; -

DR EMBL; X59720; CAA42373.1; -

DR EMBL; X54535; CAA38402.1; -

DR EMBL; D00842; BAA00723.1; -

DR PIR; JX0182; ISBYS.

DR PIR; S19372; S19372.

DR PIR; S15050; S15050.

DR PIR; S25349; S25349.

DR PIR; A41713; A41713.

DR PIR; U50634; U50634.

DR HSSP; P07237; IMEX.

DR SGD; S0000548; PDI1.

DR InterPro; IPR000886; ER target.

DR InterPro; IPR00063; Thiored.

DR Pfam; PF00085; thiored; 2.

DR PRINTS; PR00421; THIREDOXIN.

DR TIGRFAMs; TIGR01126; pdi\_dom; 2.

DR TIGRFAMs; TIGR01130; ER\_PDI\_fam; 1.

DR PROSITE; PS00014; ER\_TARGET; 1.

DR PROSITE; PS00194; THIREDOXIN; 2.

KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.

FT SIGNAL 1 28

FT CHAIN 1 522

FT DISULFID 61 522

FT DISULFID 64 64

FT DISULFID 406 409

FT CARBOHYD 82 82

FT CARBOHYD 117 117

FT CARBOHYD 155 155

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROHD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CAROHD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 519 522 PREVENT SECRETION FROM ER.  
 FT CONFLICT 33 52 L1SWPTLSTMTSFRSTW (IN REF. 3).  
 FT CONFLICT 83 83 I -> V (IN REF. 4).  
 FT CONFLICT 114 114 S -> R (IN REF. 6).  
 FT CONFLICT 143 143 V -> S (IN REF. 4).  
 FT CONFLICT 146 146 MISSING (IN REF. 4).  
 FT CONFLICT 168 168 K -> E (IN REF. 4).  
 FT CONFLICT 197 197 D -> R (IN REF. 3).  
 FT CONFLICT 215 215 A -> V (IN REF. 4).  
 FT CONFLICT 226 226 E -> S (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 333 333 S -> P (IN REF. 3).  
 FT CONFLICT 351 351 L -> F (IN REF. 3).  
 FT CONFLICT 455 455 G -> S (IN REF. 4).  
 FT CONFLICT 458 458 A -> AADMAEA (IN REF. 4 AND 6).  
 FT CONFLICT 505 505  
 SQ SEQUENCE 522 AA; 58227 MW; 69CF3B05D7F74C94 CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 522;  
 Best Local Similarity 38.5%; Pred. No. 22;  
 Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 GGNMNIKLKMPY 15  
 DB 302 AGNLMKEQFPLF 314  
 RESULT 10  
 Y344 CHLPN STANDARD; PRT; 621 AA.  
 AC Q9K275; Q9JSG0; Q9Z8J9;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical zinc metalloprotease (CP0344/CP0344 (EC 3.4.24.-)).  
 GN CP0344 OR CP0416 OR CP0344.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
 OC NCBI\_TaxID=83558;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunnham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Ulfenback T., Berry K., Bass S.,  
 RA Liner K., Weidman U., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Eisen M., Nelson W., Deboy R., Kolonyay J., McClarty G., Salzberg S.L.,  
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 CC from Japan and CWL029 from USA.";  
 CC Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- COFACTOR: Zinc (Probable).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

(By similarity). BELONGS TO PEPTIDASE FAMILY M50B.  
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; AB001618; AAD18488.1; -  
 DR EMBL; AB002203; AAF8260.1; -  
 DR EMBL; AP002546; BAA8552.1; -  
 DR TIGR; CP0416; -  
 DR InterPro; IPR001478; PDZ.  
 DR SMART; SM00228; PDZ; 2.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;  
 KW Inner membrane; Complete proteome.  
 KM Inner membrane; Complete proteome.  
 FT METAL 20 20 ZINC (CATALYTIC) (POTENTIAL).  
 FT ACT SITE 21 21 POTENTIAL.  
 FT METAL 24 24 ZINC (CATALYTIC) (POTENTIAL).  
 FT TRANSMEM 103 125 POTENTIAL.  
 FT TRANSMEM 561 583 POTENTIAL.  
 FT TRANSMEM 596 613 POTENTIAL.  
 FT CONFLICT 224 224 G -> S (IN REF. 1).  
 FT CONFLICT 584 584 T -> K (IN REF. 1).  
 FT CONFLICT 598 598 V -> D (IN REF. 3).  
 SQ SEQUENCE 621 AA; 69820 MW; 30B09776F8A271BF CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 621;  
 Best Local Similarity 33.3%; Pred. No. 27;  
 Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 IFSGNMNIKLKMPY 15  
 DB 178 LLEGHNLRIKRGY 192  
 RESULT 11  
 PTP3 YEAST STANDARD; PRT; 928 AA.  
 AC P40048;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-FEB-1998 (Rel. 37, Last annotation update)  
 DE Protein-tyrosine phosphatase 3 (EC 3.1.3.48) (PTPASE 3).  
 GN PTP3 OR YER075C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RX MEDLINE=97367951; PubMed=9224718;  
 RA Zhan X.-L., Deschenes R.D., Guan K.-L.;  
 RT "Differential regulation of FUS3 MAP kinase by tyrosine-specific  
 RT phosphatases PTP2/PTP3 and dual-specificity phosphatase MSG5 in  
 RT Saccharomyces cerevisiae.";  
 RL Genes Dev. 11:1690-1702(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
 RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
 RA Chung E., Duncan M., Guzman E., Hartzell G., Huntke-Smith S.,  
 RA Hyman R., Kayser A., Kemp C., Isakari D., Lew H., Lin D.,  
 RA Moseedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;  
 RT Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.



CC -!-FUNCTION: MAJOR PHOSPHATASE RESPONSIBLE FOR TYROSINE  
 CC DEPHOSPHORYLATION OF MAP KINASE FUS3 TO INACTIVATE ITS ACTIVITY;  
 CC IT ALSO HAS IMPORTANT ROLES, ALONG WITH MSG5, IN THE INACTIVATION  
 CC OF FUS3 FOLLOWING PHEROMONE STIMULATION.  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 CC tyrosine + phosphate  
 CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AF006304; AAB70811.1; -  
 CC DR EMBL; U18814; AAB64614.1; -  
 CC DR HSSP; P29350; IGWZ.  
 CC DR SGD; S0000877; PTP3.  
 CC DR InterPro; IPR001763; Rhodanese-like.  
 CC DR InterPro; IPR000387; TYR phosphatase.  
 CC DR Pfam; PF00102; Y.phosphatase; 1.  
 CC DR Pfam; PF00581; Rhodanese; 1.  
 CC DR SMART; SM00194; PTPC; 1.  
 CC DR SMART; SM00450; RHOD; 1.  
 CC DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 CC DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 CC DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 CC KM Hydrolyase.  
 CC FT ACT SITE 804 804 BY SIMILARITY.  
 CC FT DOMAIN 308 311 POLY-SER.  
 CC FT DOMAIN 592 597 POLY-THR.  
 CC FT DOMAIN 702 715 POLY-ASN.  
 CC FT DOMAIN 724 732 POLY-ASP.  
 CC SQ SEQUENCE 928 AA; 105250 MW; 6DECBA26B7ACB6B CRC64;  
 CC  
 CC Query Match 46.8%; Score 37; DB 1; Length 928;  
 CC Best Local Similarity 63.6%; Pred. No. 43;  
 CC Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC QY 3 SGNNMILKMP 13  
 CC Db 272 SPQNNLKLKVP 282  
 CC  
 CC RESULT 12  
 CC EPIB\_STAEP STANDARD; PRT; 966 AA.  
 CC ID EPIB\_STAEP  
 CC AC P30195;  
 CC DT 01-APR-1993 (Rel. 25, Created)  
 CC DT 01-APR-1993 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Epidermal biosynthesis protein epib.  
 CC GN EPIB.  
 CC OS Staphylococcus epidermidis.  
 CC OG Plasmid pTU 32.  
 CC OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CC OX NCBI\_TaxID=1282;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=RU 3298 / DSM 3095;  
 CC RX MEDLINE=92155237; PubMed=1740156;  
 CC RA Schell N., Engelse G., Augustin J., Rosenstein R., Ungermann V.,  
 CC Goetz F., Entian K.-D.;  
 CC RT "Analysis of genes involved in the biosynthesis of lantibiotic  
 CC epidermin."  
 CC BU: J. Biochem. 204:57-68(1992).  
 CC -!- FUNCTION: INVOLVED IN THE POSTTRANSLATIONAL MODIFICATION OF THE  
 CC LANTIBIOTIC EPIDERMIN.  
 CC CC -!- SUBCELLULAR LOCATION: POSSIBLY ASSOCIATED WITH, AND ANCHORED TO,

CC THE CYTOPLASMIC SIDE OF THE MEMBRANE.  
 CC -!- SIMILARITY: TO B.SUBTILIS SPAN AND L.LACTIS NISB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; X62386; CAA4253.1; -  
 CC KW Plasmid.  
 CC SQ SEQUENCE 986 AA; 116708 MW; 343CE98526847850 CRC64;  
 CC  
 CC Query Match 46.8%; Score 37; DB 1; Length 986;  
 CC Best Local Similarity 70.0%; Pred. No. 46;  
 CC Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC QY 2 FSGNNMILK 11  
 CC Db 462 FTGNNMILK 471  
 CC  
 CC RESULT 13  
 CC DYHC\_EMENT STANDARD; PRT; 4344 AA.  
 CC ID DYHC\_EMENT  
 CC AC P45444;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 CC DE Dynein heavy chain, cytosolic (DYHC).  
 CC GN NUDA.  
 CC OS Emericella nidulans (Aspergillus nidulans).  
 CC OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC OC Eurotiiales; Trichocomaceae; Emericella.  
 CC OX NCBI\_TaxID=5072;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=94181539; PubMed=8134356;  
 CC RA Xiang X., Beckwith S.M., Morris R.N.;  
 CC RT "Cytoplasmic dynein is involved in nuclear migration in Aspergillus  
 CC nidulans."  
 CC RT Proc. Natl. Acad. Sci. U.S.A. 91:2100-2104(1994).  
 CC CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
 CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
 CC ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
 CC NUCLEAR DISTRIBUTION IN HYPHAE.  
 CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
 CC INTERMEDIATE AND LIGHT CHAINS.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U03904; AAI18338.1; -  
 CC DR PIR; A53489; A53489.  
 CC DR InterPro; IPR004273; Dynein heavy.  
 CC DR Pfam; PF03028; Dynein heavy; 1.  
 CC KW Motor protein; Microtubules;  
 CC FT DOMAIN 667 687 Dynein; ATP-binding, Coiled coil.  
 CC FT DOMAIN 913 933 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1321 1341 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1548 1564 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1627 1656 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 2035 2063 MICROTUBULE-BINDING (POTENTIAL).  
 CC FT DOMAIN 3179 3270 COILED COIL (POTENTIAL).

FT DOMAIN 3407 3463 COILED COIL (POTENTIAL).  
 FT DOMAIN 3700 3798 COILED COIL (POTENTIAL).  
 FT NP BIND 1933 1940 ATP (POTENTIAL).  
 FT NP BIND 2223 2230 ATP (POTENTIAL).  
 FT NP BIND 2592 2599 ATP (POTENTIAL).  
 FT NP BIND 2932 2939 ATP (POTENTIAL).  
 SQ SEQUENCE 4344 AA; 492470 MW; 302A588CAF8BD019 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 4344;  
 Best Local Similarity 42.9%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 IFSGNNMIKMKPM 14  
 DB 1584 VFTGNADIKILPV 1597

RESULT 14  
 DYHC DROME STANDARD; PRT; 4639 AA.  
 AC 37276;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Dynein heavy chain, cytosolic (DYHC).  
 OS CDHC OR DHC64C.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94375524; PubMed=8089180;  
 RA Li M., McGrail M., Serr M., Hays T.S.;  
 RT "Drosophila cytoplasmic dynein, a microtubule motor that is  
 RT asymmetrically localized in the oocyte."  
 RL J. Cell Biol. 126:1475-1494(1994).  
 [2]  
 RN SEQUENCE OF 1877-1998 FROM N.A.  
 RX MEDLINE=94243034; PubMed=8186464;  
 RA Rammason K., Serr M., Gerner J., Gibbons I., Hays T.S.;  
 RT "A family of dynein genes in Drosophila melanogaster."  
 RL Mol. Biol. Cell 5:45-55(1994).  
 CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
 CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
 CC ORGANELLES ALONG MICROTUBULES.  
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
 CC INTERMEDIATE AND LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L23195; AAA60323.1; -  
 CC EMBL: L25122; AAA28482.1; -  
 CC FlyBase; FBgn010349; Dnc64C.  
 DR InterPro; IPR004273; Dynein\_heavy.  
 DR Pfam; PF03028; Dynein\_heavy; 1.  
 KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.  
 FT DOMAIN 530 565 COILED COIL (POTENTIAL).  
 FT DOMAIN 774 794 COILED COIL (POTENTIAL).  
 FT DOMAIN 1264 1368 COILED COIL (POTENTIAL).  
 FT DOMAIN 1999 2027 COILED COIL (POTENTIAL).  
 FT DOMAIN 3189 3261 MICROTUBULE-BINDING (POTENTIAL).  
 FT DOMAIN 3382 3478 COILED COIL (POTENTIAL).  
 FT DOMAIN 3723 3782 COILED COIL (POTENTIAL).

FT NP BIND 1895 1902 ATP (POTENTIAL).  
 FT NP BIND 2210 2217 ATP (POTENTIAL).  
 FT NP BIND 2580 2587 ATP (POTENTIAL).  
 FT NP BIND 2922 2929 ATP (POTENTIAL).  
 SQ SEQUENCE 4639 AA; 530152 MW; 057A7D8800CCDD07E CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 4639;  
 Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 IFSGNNMIKMKPM 14  
 DB 1540 IFSGSADIKILPV 1553

RESULT 15  
 TR33 YEAST STANDARD; PRT; 268 AA.  
 AC 099394;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Transport protein particle 33 kDa subunit (TRAP 33 kDa subunit).  
 GN TR33 OR YOR115C OR O3251 OR YOR3251C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 NCBI\_Taxid=4932;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=S288C / FY1679;  
 RC MEDLINE=97060020; PubMed=8904341;  
 RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vlcek C.,  
 RA Stegemann U., Zimmernann J., Bfelle H., Paces V., Ansoerge W.;  
 RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV  
 RT from Saccharomyces cerevisiae reveals 30 open reading frames."  
 RL Yeast 12:281-288(1996).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=97344368; PubMed=9200815;  
 RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,  
 RA Schwager C., Paces V., Sander C., Ansoerge W.;  
 RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV."  
 RL Yeast 13:655-672(1997).  
 CC -1- FUNCTION: TRAP plays a key role in the late stages of endoplasmic  
 CC reticulum to Golgi traffic.  
 CC -1- SUBUNIT: Part of multisubunit complex composed of BET3, BET5,  
 CC TRS20, TRS23, TRS31, TRS33, TRS65, TRS85, TRS120 and TRS130.  
 CC -1- SUBCELLULAR LOCATION: cis-Golgi complex.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X90518; CAA62124.1; -  
 CC DR EMBL: X94335; CAA64035.1; -  
 CC DR EMBL: Z75023; CAA99313.1; -  
 CC SCD; S0005641; TRS33.  
 KW Transport; Endoplasmic reticulum; Golgi stack.  
 SQ SEQUENCE 268 AA; 30749 MW; 2480F69D7A4ED363 CRC64;

Query Match 45.6%; Score 36; DB 1; Length 268;  
 Best Local Similarity 63.6%; Pred. No. 16;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 IFSGNNMIKMK 11  
 DB 139 IFSNPNLKF 149

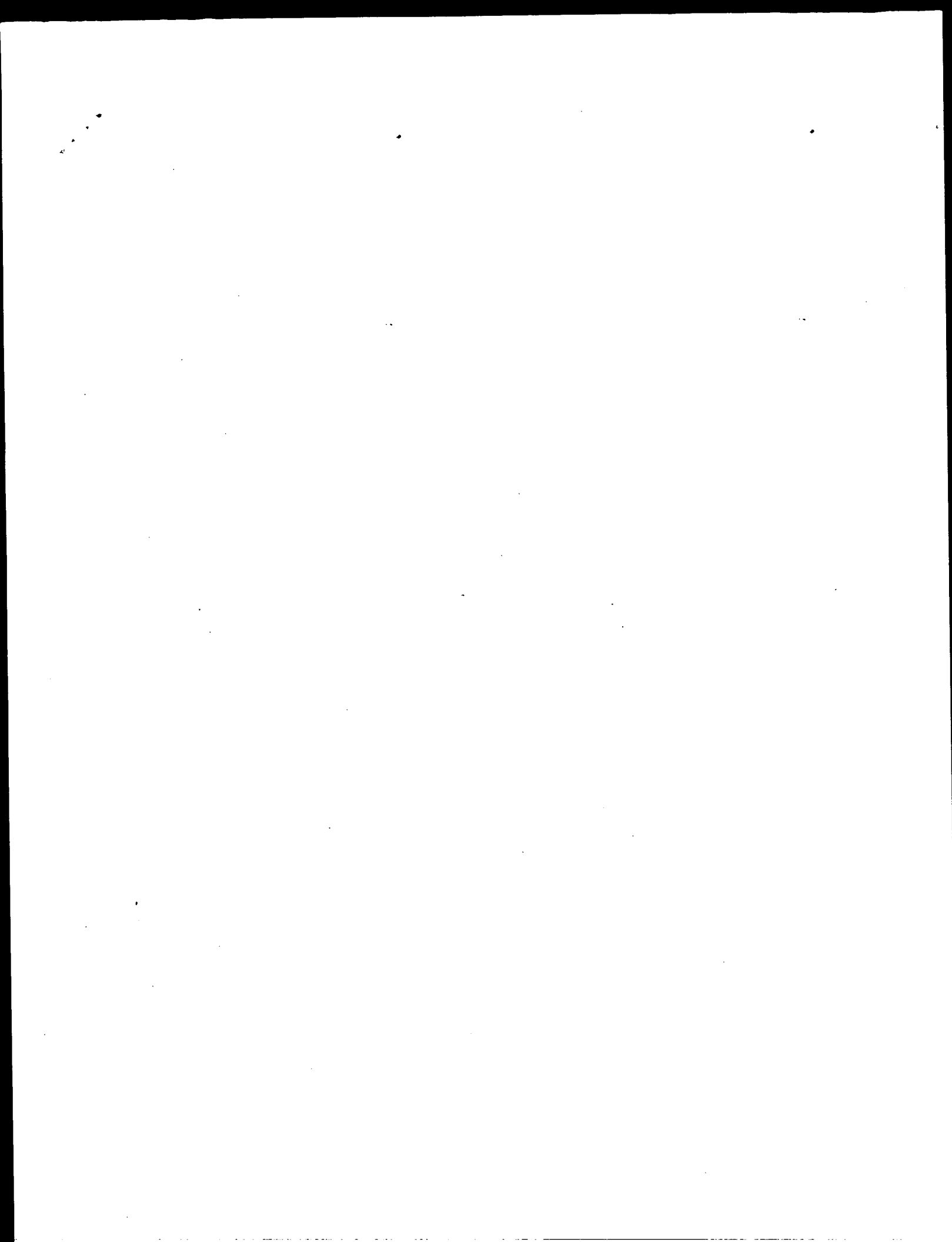
Sat May 10 08:44:08 2003

us-09-142-524d-29.rsp

Page 9

Search completed: April 20, 2003, 13:07:32  
Job time : 5.92105 secs

---



GenCore version 5.1.4 p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds  
(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524D-29

Perfect score: 79

Sequence: 1 IFSGNNMIKMKMPY 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_21.\*  
2: sp\_archaea.\*  
3: sp\_bacteria.\*  
4: sp\_fungi.\*  
5: sp\_human.\*  
6: sp\_invertebrate.\*  
7: sp\_mammal.\*  
8: sp\_mhc.\*  
9: sp\_organelle.\*  
10: sp\_phage.\*  
11: sp\_plant.\*  
12: sp\_rodent.\*  
13: sp\_virus.\*  
14: sp\_vertebrate.\*  
15: sp\_unclassified.\*  
16: sp\_rvitus.\*  
17: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	374	10 Q8RUR1	Q8RUR1 cryptomeria
2	68	86.1	367	10 Q9LIT2	Q9LIT2 juniperus v
3	68	86.1	367	10 Q9LIT1	Q9LIT1 juniperus v
4	68	86.1	367	10 Q9M4S6	Q9M4S6 cupressus s
5	68	86.1	367	10 Q9M4S4	Q9M4S4 cupressus s
6	68	86.1	367	10 Q9M4S3	Q9M4S3 cupressus s
7	68	86.1	367	10 Q9M4S2	Q9M4S2 cupressus s
8	68	86.1	367	10 Q9M4S1	Q9M4S1 juniperus o
9	64	81.0	367	10 Q9M4S5	Q9M4S5 cupressus s
10	64	81.0	367	10 Q9M4S6	Q9M4S6 cupressus s
11	42	53.2	272	8 Q9B295	Q9B295 ceratopsolen
12	42	53.2	312	11 Q8VG18	Q8VG18 mus musculu
13	42	53.2	521	8 Q9B975	Q9B975 ceratopsolen
14	42	53.2	543	12 Q83893	Q83893 ovine adeno
15	40	50.6	245	12 Q84447	Q84447 paramacium
16	40	50.6	262	8 Q21044	Q21044 dictyostell

17	40	50.6	341	16 Q928Q8	Q928Q8 chlamydia p
18	40	50.6	445	9 Q8SD05	Q8SD05 pseudomonas
19	40	50.6	578	16 Q93X95	Q93X95 staphylococ
20	40	50.6	888	5 Q95U80	Q95U80 drosophila
21	40	50.6	1069	5 Q9VF64	Q9VF64 drosophila
22	40	50.6	1084	5 Q8SWR8	Q8SWR8 drosophila
23	39	49.4	306	2 Q9L973	Q9L973 lactococcus
24	39	49.4	376	16 Q9CMT9	Q9CMT9 pasteurella
25	39	49.4	421	11 Q9OYR8	Q9OYR8 mus musculu
26	39	49.4	766	5 Q25823	Q25823 plasmodium
27	39	49.4	794	16 Q8XB12	Q8XB12 escherichia
28	39	49.4	1611	16 Q8RFV3	Q8RFV3 fusobacteri
29	38	48.1	272	8 Q9G7X0	Q9G7X0 meselatus f
30	38	48.1	283	17 Q97YF1	Q97YF1 sulfolobus
31	38	48.1	304	2 Q9EVT5	Q9EVT5 thermotoga
32	38	48.1	310	2 Q93L00	Q93L00 bacillus th
33	38	48.1	316	16 Q8RHM2	Q8RHM2 fusobacteri
34	38	48.1	326	11 Q8VFR1	Q8VFR1 mus musculu
35	38	48.1	347	8 Q8RNV6	Q8RNV6 guillardia
36	38	48.1	369	2 Q8RSW6	Q8RSW6 synchococc
37	38	48.1	371	10 Q94FP0	Q94FP0 arabidopsis
38	38	48.1	440	12 Q9EMV0	Q9EMV0 amesacea moo
39	38	48.1	442	2 Q9KX32	Q9KX32 streptococc
40	38	48.1	488	10 Q22962	Q22962 arabidopsis
41	38	48.1	501	16 Q8RBU7	Q8RBU7 thermoaer
42	38	48.1	757	5 Q21407	Q21407 caenorhabdi
43	38	48.1	992	11 Q9WU50	Q9WU50 mus musculu
44	38	48.1	1007	5 Q9U0D1	Q9U0D1 aplysia cal
45	38	48.1	1344	11 Q35851	Q35851 mus musculu

## ALIGNMENTS

Q8RUR1	PRELIMINARY;	PRT;	374 AA.
AC Q8RUR1;			
DT 01-JUN-2002 (TREMUR1. 21, Created)			
DT 01-JUN-2002 (TREMUR1. 21, Last sequence update)			
DT 01-JUN-2002 (TREMUR1. 21, Last annotation update)			
DE Cry j 1 precursor.			
GN Cry j 1.1 OR Cry j 1.2.			
OS Cryptomeria japonica (Japanese cedar).			
OC Embryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.			
OX NCBI_TaxID=3369;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TRISUB=POLLEN;			
RA Futamura N., Shinohara K.;			
RT "Isolation and characterization of cDNAs encoding major allergen Cry j			
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AB081309; BAB86286.1; -			
DR EMBL; AB081310; BAB86287.1; -			
KM Signal.			
FT SIGNAL	1	21	POTENTIAL.
FT CHAIN	22	374	Cry j 1.
SQ SEQUENCE	374 AA;	40720 MW;	90D0085D24BF2BD4 CRC64;
Query Match	100.0%;	Score 79;	DB 10;
Best Local Similarity	100.0%;	Pred. No. 2.7e-06;	
Matches	15;	Conservative	0;
		Mismatches	0;
		Indels	0;
		Gaps	0;
QY	1 IFSGNNMIKMKMPY 15		
DB	92 IFSGNNMIKMKMPY 106		
RESULT 2			
Q9LIT2	PRELIMINARY;	PRT;	367 AA.
ID Q9LIT2			

AC 09LTL2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-2.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=39584;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21315424; PubMed=11422137;  
 RX Midoro-Horvut T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Exp. Allergy 31:771-778(2001).  
 DR EMBL; AF151427; AAF80164.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39768 MW; 0A6AC2F1BAF69586 CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

RESULT 3  
 ID 09LTL1 PRELIMINARY; PRT; 367 AA.  
 AC 09LTL1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-1.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=39584;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21315424; PubMed=11422137;  
 RX Midoro-Horvut T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Exp. Allergy 31:771-778(2001).  
 DR EMBL; AF151429; AAF80166.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39708 MW; DCBD1981A74E4711 CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

RESULT 4  
 ID 09M4S6 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 Pollen allergen.

GN CUP51.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF257491; AAF72625.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39835 MW; BA9CC0108C2C5A3 CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

RESULT 5  
 ID 09M4S4 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S4;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP51.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF257493; AAF72627.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39934 MW; 97AD3011D74E3D6E CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

RESULT 6  
 ID 09M4S3 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S3;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP51.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monesalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39832 MW; B5DBPFA61C07A53 CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

## RESULT 7

ID Q9M4S2 PRELIMINARY; PRT; 367 AA.  
 AC Q9M4S2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSL.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monesalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257495; AAF72629.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39819 MW; AE7E055A61C07A53 CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

## RESULT 8

ID Q93X51 PRELIMINARY; PRT; 367 AA.  
 AC Q93X51;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Putative allergen Jun o 1.  
 GN JUN O 1.  
 OS Juniperus oxycedrus (Prickly Juniper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxID=99008;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=POLLEN;  
 RC TISSUE=POLLEN;  
 FT TISSUE=POLLEN;  
 RT "Cloning of Juniperus oxycedrus major allergen.";

RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ293767; CAC48400.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 SQ SEQUENCE 367 AA; 39808 MW; 5D28204DBFDB19D7 CRC64;

Query Match 86.1%; Score 68; DB 10; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 0.00028;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

## RESULT 9

ID Q9M4S5 PRELIMINARY; PRT; 367 AA.  
 AC Q9M4S5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSL.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monesalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens."  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257492; AAF72626.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39894 MW; 5D56FC0E3263B741 CRC64;

Query Match 81.0%; Score 64; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.0015;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

## RESULT 10

ID Q93XL6 PRELIMINARY; PRT; 367 AA.  
 AC Q93XL6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative allergen Cup a 1 precursor.  
 GN CUP A 1.  
 OS Cupressus arizonica.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=49011;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=POLLEN;  
 RC TISSUE=POLLEN;  
 FT TISSUE=POLLEN;  
 RT "Cloning of Cupressus arizonica major allergen."  
 RL Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ278498; CAC37790.2; -  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN.  
 SQ SEQUENCE 367 AA; 39809 MW; AFP97260423A9F28 CRC64;

Query Match 81.0%; Score 64; DB 10; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.0015;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IFSGNNIKLKMPLY 15  
 Db 92 IFSGNNIKLKMPLY 106

## RESULT 11

Q9B295 PRELIMINARY; PRT; 272 AA.  
 AC Q9B295;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).  
 GN Cyt.  
 OS Ceratopsolen bisulcatus.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;  
 OC Chalcidoidea; Agaonidae; Agaoninae; Ceratopsolen.  
 OX NCBI\_TaxID=130004;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C.BISULC;  
 RX MEDLINE=21217342; PubMed=11321056;  
 RA Machado C.A., Joussetin B., Kjellberg F., Compton S.G., Herre E.A.;  
 RT "Phylogenetic relationships, historical biogeography, and character  
 evolution of fig pollinating wasps."  
 RL Proc. R. Soc. Lond. B, Biol. Sci. 268:685-694 (2001).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 C + 2 H(2)O.  
 CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC -1- INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL; AY014986; AAK13227.1; -.  
 DR InterPro; IPR000883; COX1.  
 DR Pfam; PF00115; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KM Respiratory chain; Transmembrane.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 272 AA; 30878 MW; E78E6C72148A1867 CRC64;

Query Match 53.2%; Score 42; DB 8; Length 272;  
 Best Local Similarity 64.3%; Pred. No. 13;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IFSGNNIKLKMPLY 14  
 Db 242 IFSGNNIKLKMPLY 255

## RESULT 12

Q8VG18 PRELIMINARY; PRT; 312 AA.  
 AC Q8VG18;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Olfactory receptor MOR112-2.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhang X., Firestein S.J.;  
 RT "The olfactory receptor gene superfamily of the mouse."  
 RL Nat. Neurosci. 0:0-0 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Adams M.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY073160; AAL60823.1; -.  
 DR InterPro; IPR000276; GPCR\_Rhodopsn.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHOPOPSN.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; UNKNOWN\_1.  
 DR PROSITE; PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
 KW Receptor.  
 SQ SEQUENCE 312 AA; 35053 MW; 80D36557ED2A01B9 CRC64;

Query Match 53.2%; Score 42; DB 11; Length 312;  
 Best Local Similarity 47.6%; Pred. No. 15;  
 Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

QY 3 SGNNITTLNLSVHAKTPMY 15  
 Db 38 SGNNITTLNLSVHAKTPMY 58

## RESULT 13

Q9B975 PRELIMINARY; PRT; 521 AA.  
 AC Q9B975;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).  
 OS Ceratopsolen bisulcatus.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;  
 OC Chalcidoidea; Agaonidae; Agaoninae; Ceratopsolen.  
 OX NCBI\_TaxID=130004;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B170;  
 RA Weiblen G.D.;  
 RT "Phylogenetic analyses of dioecious fig pollinators based on  
 mitochondrial DNA sequences and morphology."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 C + 2 H(2)O.  
 CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC -1- INNER MEMBRANE (BY SIMILARITY).  
 CC EMBL; AF200375; AAK0062.1; -.  
 DR HSRP; P18401; 1PPT.  
 DR InterPro; IPR000883; COX1.  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASE1.  
 DR PROSITE; PS00077; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;



KM Respiratory chain; Transmembrane.  
 PT NON-TER 1  
 SQ SEQUENCE 521 AA; 58771 MW; 164C7EF2C4B66C5B CRC64;

Query Match 53.2%; Score 42; DB 8; Length 521;  
 Best Local Similarity 64.3%; Pred. No. 24;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IFSGNNMNIKLKMPM 14  
 DB 464 IFSGNNMNIKLKMPM 477

## RESULT 14

ID Q83893 PRELIMINARY; PRT; 543 AA.

AC Q83893; 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE Fiber.  
 OS Ovine adenovirus type 7.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Atadenovirus.

NCBI\_TaxID=114430;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=OAV287;  
 RX MEDLINE=95297141; PubMed=7778275;  
 RA Virat S., Boyle D., Kocherhans R., Both G.W.;

RT "Sequence of ovine adenovirus homologs for 100K hexon assembly, 33K, pVIII, and fiber genes: early region E3 is not in the expected location";  
 RL Virology 209:400-408(1995).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OAV287;

RX MEDLINE=96240641; PubMed=8659111;  
 RA Virat S., Brookes D.E., Strike P., Khatri A., Boyle D.B., Both G.W.;

RT "Unique genome arrangement of an ovine adenovirus: identification of new proteins and proteinase cleavage sites";  
 RL Virology 220:186-199(1996).

RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OAV287;

RX MEDLINE=97271311; PubMed=9126262;  
 RA Xu Z.Z., Hyatt A., Boyle D.B., Both G.W.;

RT "Construction of ovine adenovirus recombinants by gene insertion or deletion of related terminal region sequences";  
 RL Virology 230:62-71(1997).

RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OAV287;  
 RX MEDLINE=98277079; PubMed=9614874;

RA Khatri A., Both G.W.;

RT "Identification of transcripts and promoter regions of ovine adenovirus OAV287";  
 RL Virology 245:128-141(1998).  
 DR EMBL: U40839; AAA73919.1; -  
 DR InterPro: IPR000939; Adeno\_fiber2.  
 DR Pfam: PF00608; adeno\_fiber2; 10.

SQ SEQUENCE 543 AA; 58217 MW; 20FA2F5CCTA8C915 CRC64;

Query Match 53.2%; Score 42; DB 12; Length 543;  
 Best Local Similarity 72.7%; Pred. No. 25;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNNMNIKLKMPM 14  
 DB 178 GNNMNIKLKMPM 188

RESULT 15  
 Q84447

ID Q84447 PRELIMINARY; PRT; 245 AA.

AC Q84447; 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE A127R protein.  
 OS Parametium bursaria chlorella virus 1 (PBCV-1).  
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.

NCBI\_TaxID=10506;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=95407089; PubMed=7676624;  
 RA Li Y., Lu Z., Burbank D.E., Kutish G.F., Rock D.L., Van Etten J.L.;

RT "Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map positions 45 to 88";  
 RL Virology 212:134-150(1995).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20013326; PubMed=10544099;

RA Kaiser A., Volmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,  
 RA Lisee A.D., Nickerson K.W., Van Etten J.L.;

RT "Chlorella virus PBCV-1 encodes a functional homospentidine synthase";  
 RL Virology 263:254-262(1999).

RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20478054; PubMed=11021991;

RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;

RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus PBCV-1";  
 RL Virology 276:27-36(2000).

RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Van Etten J.L.;

RT Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.

RA Van Etten J.L.;

RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.

RA Graves M.V., Van Etten J.L.;

RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.

RA Graves M.V., Van Etten J.L.;

RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE FROM N.A.

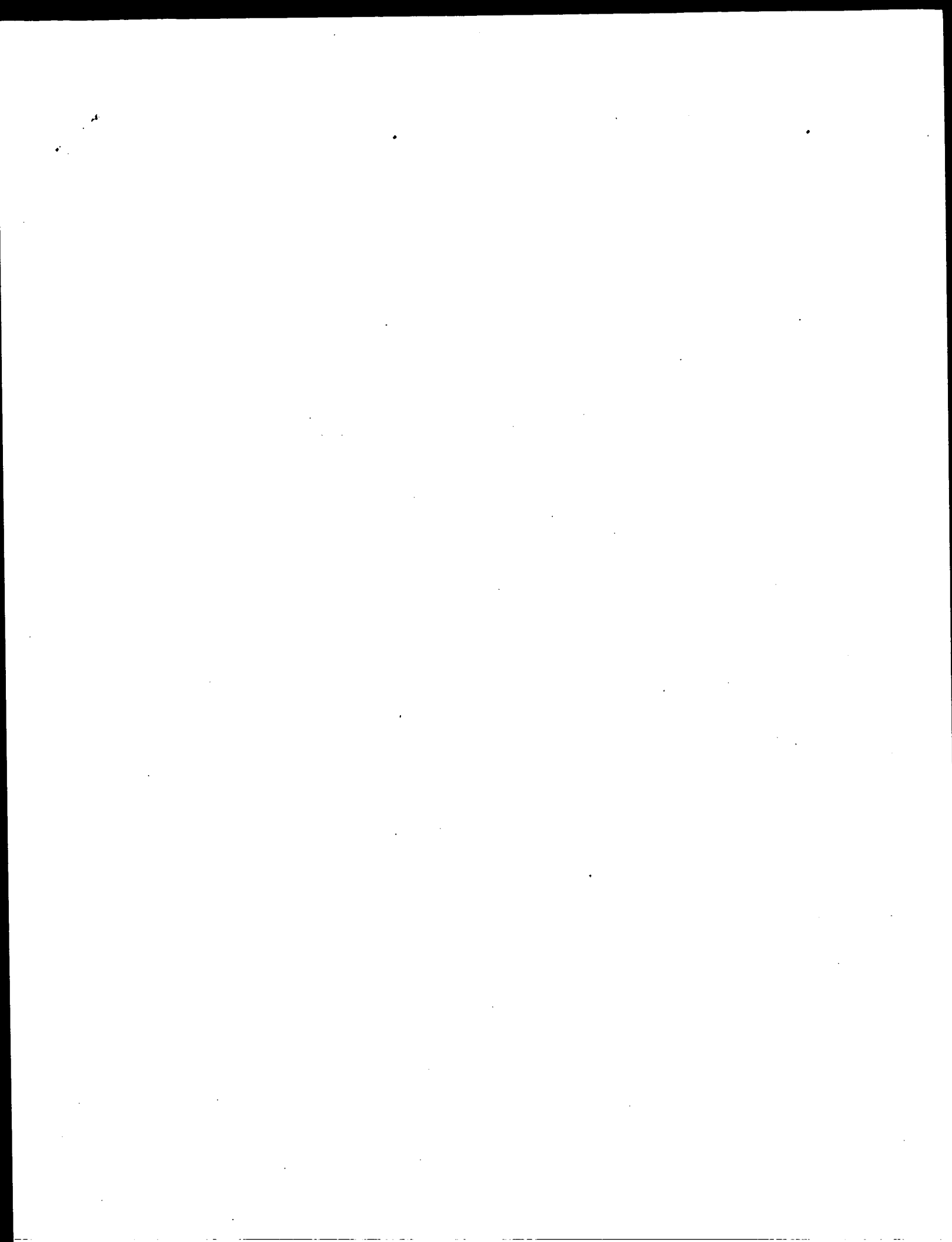
RA Gurnon J.R., Graves M.V., Van Etten J.L.;

RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U42580; AAC96495.1; -  
 SQ SEQUENCE 245 AA; 27126 MW; C8DC76A169CDE6 CRC64;

Query Match 50.6%; Score 40; DB 12; Length 245;  
 Best Local Similarity 72.7%; Pred. No. 27;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 FSGNNMNIKLKMPM 12  
 DB 62 FSGNNMNIKLKMPM 72

Search completed: April 20, 2003, 13:13:01  
 Job time : 14.6711 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-36

Sequence: 1 PCVFIKRVSNVTHG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	374	1 SBP_CRYUA	P18632 cryptomeria
2	58	72.5	375	1 MPAL_CHAOB	O96385 chamaecypar
3	57	71.2	346	1 MPAL_CUPAR	O96399 cupressus a
4	55	68.8	367	1 MPAL_TUNAS	P81294 juniperus a
5	55	68.8	404	1 9612_LYCS	P24396 lycopersico
6	42	52.5	290	1 KEPR_NITVU	P37100 nitrobacter
7	41	51.2	230	1 SFSA_PYRVU	O8u18 pyrococcus
8	39	48.8	439	1 LIPI_DROME	O46107 dirosophila
9	39	48.8	720	1 VMS2_YEAST	O04322 saccharomyc
10	38	47.5	158	1 PRPB_YEAST	P42909 escherichia
11	38	47.5	214	1 Y023_CAEEL	P34673 caenorhabdi
12	38	47.5	587	1 REL_MOUSE	P15307 mus musculu
13	38	47.5	1328	1 POLX_TOBAC	P10978 nicotiana t
14	37	46.2	146	1 SP22_BACSH	O32724 bacillus sp
15	37	46.2	290	1 ATNC_HUMAN	P14231 mus musculu
16	37	46.2	290	1 ATNC_MOUSE	P13638 rattus norv
17	37	46.2	290	1 ATNC_MOUSE	P13638 rattus norv
18	37	46.2	315	1 TRXB_MYGE	P43348 mycoplasma
19	37	46.2	326	1 PEL_EMENT	O00645 emericella
20	37	46.2	449	1 PEL_YEAST	P15722 lycopersico
21	37	46.2	452	1 IE63_VZVD	P09269 varicella-z
22	37	46.2	722	1 VGLH_GPCMV	P87730 guinea pig
23	37	46.2	985	1 NAHI_YEAST	O98271 saccharomyc
24	37	46.2	4128	1 PRKD_MOUSE	P97313 mus musculu
25	36	45.0	124	1 FOLB_CHLTR	O84620 chlamydia t
26	36	45.0	124	1 FOLB_CHLTR	O84620 chlamydia t
27	36	45.0	434	1 HIS8_CAMCE	O96112 campylobact
28	36	45.0	488	1 EXON_HSV6Z	P40973 human long
29	36	45.0	601	1 TCPE_YEAST	P54413 human herpe
30	36	45.0	601	1 TCPE_YEAST	P54413 human herpe
31	36	45.0	688	1 Y080_SCHPO	O18803 schistosom
32	36	45.0	914	1 NRPI_CHICK	P47147 saccharomyc
33	36	45.0	914	1 NRPI_CHICK	P47147 saccharomyc
					P79795 gallus galli
					P35220 dirosophila

34	36	45.0	922	1 NRPI_RAT	O96419 rattus norv
35	36	45.0	923	1 NRPI_HUMAN	O14786 homo sapien
36	36	45.0	923	1 NRPI_MOUSE	P97333 mus musculu
37	35.5	44.4	350	1 PYRC_BUCAI	P57416 buchnera ap
38	35	43.8	123	1 FERS_RHOCA	P37097 rhodobacter
39	35	43.8	164	1 MCRW_META	O58253 methanococc
40	35	43.8	183	1 Y584_CHLTR	O84588 chlamydia t
41	35	43.8	183	1 Y584_CHLTR	O84588 chlamydia t
42	35	43.8	290	1 ATNC_BOVIN	O96116 chlamydia m
43	35	43.8	317	1 Y640_AQUAE	O28030 bos taurus
44	35	43.8	374	1 Y0K7_CAEEL	O66883 aquifex aeo
45	35	43.8	379	1 COS2_YEAST	O09289 caenorhabdi
					P38363 saccharomyc

## ALIGNMENTS

RESULT 1  
ID SBP\_CRYUA STANDARD; PRT: 374 AA.  
AC P18632;  
DT 01-NOV-1990 (Rel. 16, Created)  
DE 01-NOV-1995 (Rel. 32, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Sugi basic protein precursor (SBP) (Major allergen Cry j 1) (Cry j 1).  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferales; Cupressaceae; Cryptomeria.  
NCBI\_TaxId=3369;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Pollen;  
RX MEDLINE=94183234; PubMed=8135802;  
RA Sone T., Komiyama N., Shimizu K., Kusakabe T., Morikubo K.,  
RA Kuno K.;  
RT "Cloning and sequencing of cDNA coding for Cry j I, a major allergen  
of Japanese cedar pollen."  
RL Biochem. Biophys. Res. Commun. 199;619-625(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pollen;  
RX MEDLINE=95003748; PubMed=7920021;  
RA Hijioka A., Matsumoto I., Kojima K., Ogawa H.;  
RA "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
(Cryptomeria japonica) pollen allergen, Cry j I."  
RL Int. Arch. Allergy Immunol. 105;196-202(1994).  
RN [5]  
RP STRUCTURE OF CARBOHYDRATES.  
RC TISSUE=Pollen;  
RX MEDLINE=95332249; PubMed=7608114;  
RA Hino K., Yamamoto S., Sano O., Taniguchi Y., Kohno K., Usui M.,  
RA "Carbohydrate structures of the glycoprotein allergen Cry j I from  
Japanese cedar (Cryptomeria japonica) pollen."  
RL J. Biochem. 117;289-295(1995).  
CC -1- PTM: CONTAINS FUCOSE/XLOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
CC -1- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR  
CC POLLEN. THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
CC -1- MISCELLANEOUS: THE SEQUENCE OF CRY j I FORM A IS SHOWN HERE. FORM

```

CC the European Bioinformatics Institute. There are no restrictions on the use of the data for any purpose, provided that the data is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
-----
CC DR EMBL; D45404; BAA08246.1; -.
CC DR InterPro; IPR002022; Amb_allergen.
CC DR Pfam; PF00544; pec_lyase; 1.
CC DR PRINTS; PR00807; AMBALLERGEN.
CC Allergen; Glycoprotein; signal.
CC K4 SIGNAL
CC FT CHAIN 1 21
CC FT CARBOHYD 22 375
CC FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 375 AA; 40258 MW; 81CD91DF7066DBBF CRC64;
-----
QY Query Match 72.5%; Score 58; DB 1; Length 375;
Best Local Similarity 60.0%; Pred. No. 0.006;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 PCVFIKYSNVITNG 15
Db 127 PCLFMRTVSHVILHG 141
-----
RESULT 3
ID MPAI_CUPAR STANDARD; PRT; 346 AA.
AC 09SGG9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DT Major pollen allergen Cup a 1.
OS Cupressus arizonica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.
OC NCBI_TaxId=9011;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20571526; PubMed=11122214;
RA Acetunha B., Del Pozo V., Minguez A., Arrieta I., Cortegano I.,
RA Cardaba E., Gallardo S., Rojo M., Palomino P., Iñabiz C.;
RT "Molecular cloning of major allergen from Cupressus arizonica pollen: Cup a 1."
RT Clin. Exp. Allergy 30:1750-1758(2000).
CC -1 SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC AMB A I/AMB A II/CRY J I SUBFAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC CC
CC EMBL; AJ243570; CAB62551.1; -.
CC InterPro; IPR002022; Amb_allergen.
CC Pfam; PF00544; pec_lyase; 1.
CC DR PRINTS; PR00807; AMBALLERGEN.
CC Allergen; Glycoprotein.
CC K4
CC FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 346 AA; 37589 MW; F181DDCADIDSDFD0 CRC64;
-----
QY Query Match 71.2%; Score 57; DB 1; Length 346;
Best Local Similarity 53.3%; Pred. No. 0.0084;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 PCVFIKRVSNVTHG 15  
 DB 106 PCLFMRKASHVILHG 120

## RESULT 4

MPAL\_JUNAS STANDARD; PRT; 367 AA.  
 ID MPAL\_JUNAS  
 AC P81294; Q9ZNU7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 OS Major pollen allergen Jun a 1 precursor.  
 DE Juniperus ashei (Ozark white cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=13101;

SEQUENCE FROM N.A., AND SEQUENCE OF 31-38; 42-50; 58-80; 88-94;  
 RP 117-124; 134-140; 160-164; 256-263 AND 322-325.  
 RC TISSUE=Pollen;  
 RX MEDLINE=99414163; PubMed=10482836;  
 RA Midozo-Horintu T.M., Goldblum R.M., Kurosky A., Wood T.G.,  
 RA Brooks E.G.;  
 RT "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 RT allergen, Jun a 1."  
 RL J. Allergy Clin. Immunol. 104:613-617(1999).  
 RN [2]

SEQUENCE OF 22-50.  
 RP TISSUE=Pollen;  
 RX MEDLINE=99414162; PubMed=10482835;  
 RA Midozo-Horintu T., Goldblum R.M., Kurosky A., Goetz D.W.,  
 RA Brooks E.G.;  
 RT "Isolation and characterization of the mountain cedar (Juniperus  
 RT ashei) pollen major allergen, Jun a 1."  
 RL J. Allergy Clin. Immunol. 104:608-612(1999).  
 CC -1- DISEASE: THIS PROTEIN CAUSES SEVERAL SEASONAL ALLERGIC RHINITIS  
 CC IN NORTH AMERICA.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF106663; AAD03609.1; -  
 DR EMBL; AF106662; AAD03608.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Allergen; Glycoprotein; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 367 MAJOR POLLEN ALLERGEN JUN A 1.  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 367 AA; 39824 MW; FC9B81E675662E49 CRC64;

Query Match 68.8%; Score 55; DB 1; Length 367;  
 Best Local Similarity 57.1%; Pred. No. 0.02;  
 Matches 8; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIKRVSNVTHG 14  
 DB 127 PCLFMRKASHVILHG 140

RESULT 5  
 9612\_LYCES

ID 9612 LYCES STANDARD; PRT; 404 AA.  
 AC P24396;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-JUN-1999 (Rel. 38, Last annotation update)  
 DE Style development-specific protein 9612 precursor.  
 GN 9612.

OS Lycopersicon esculentum (Tomato).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Assteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 NCBI\_TaxID=4081;  
 RN [1]

SEQUENCE FROM N.A.  
 RP STRAIN=cv. VF36; TISSUE=Pistil;  
 RX MEDLINE=91117185; PubMed=2277637;  
 RA Budellier K.A., Smith A.G., Gasser C.S.;  
 RT "Regulation of a stylar transmitting tissue-specific gene in  
 RT wild-type and transgenic tomato and tobacco."  
 RL Mol. Gen. Genet. 224:183-192(1990).  
 CC -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPMENT OF THE TRANSMITTING  
 CC TISSUE OF THE STYLE AND/OR IN THE EVENTS RELATED TO POLLINATION  
 CC SUCH AS SOME ASPECT IN THE FACILITATION OF COMPATIBLE POLLEN TUBE  
 CC GROWTH.

CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN THE PISTIL WHERE IT IS  
 CC FOUND IN THE OUTER FIVE LAYERS OF THE STAMENS OF TRANSMITTING  
 CC TISSUE WITHIN THE UPPER TWO-THIRDS OF THE STYLE. FOUND AT MUCH  
 CC LOWER LEVELS IN THE ANTERS AND VEGETATIVE ORGANS.  
 CC -1- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND DURING ANTHERESIS.  
 CC -1- SIMILARITY: 54% IDENTICAL TO TOMATO PROTEINS P59 (AC P15722)  
 CC AND P56 (AC P15721).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X55193; CAA38979.1; -  
 DR PIR; S12209; S12209.  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 404  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 404 AA; 44298 MW; B2CED6B8128D8675 CRC64;

Query Match 68.8%; Score 55; DB 1; Length 404;  
 Best Local Similarity 60.0%; Pred. No. 0.022;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 PCVFIKRVSNVTHG 15  
 DB 158 PCLTHHTSNVTHG 172

RESULT 6  
 KPPR\_NITVU STANDARD; PRT; 290 AA.  
 AC P37100;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phosphoribulokinase (EC 2.7.1.19) (Phosphopentokinase) (PRK).  
 GN CHBP  
 OS Nitrobacter vulgaris.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Bradyrhizobium group; Nitrobacteraceae; Nitrobacter.  
 OX NCBI\_TaxID=29421;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=73;  
 RA Strecker M., Sickinger E., English R.S., Shively J.M., Bock E.;  
 RL Submitted (XXY-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-ribulose 5-phosphate = ADP + D-ribulose 1,5-bisphosphate.  
 CC -1- PATHWAY: Calvin cycle.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHORIBULOKINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L22884; AAA25506.1; -  
 DR HSSP; P12033; 1A7G.  
 DR InterPro; IPR001324; PRK.  
 DR Pfam; PF00485; PRK; 1.  
 DR PRINTS; PR00478; PHOSPHORIBULOKINASE.  
 DR PROSITE; PS00567; PHOSPHORIBULOKINASE; 1.  
 DR Transferrase; Kinase; Calvin cycle; ATP-binding; Photosynthesis.  
 KW NP BIND 12  
 FT ATP (POTENTIAL).  
 SQ SEQUENCE 290 AA; 32908 MW; 9E2972219A2B41C CRC64;  
 Query Match 52.5%; Score 42; DB 1; Length 290;  
 Best Local Similarity 40.0%; Pred. No. 3.6;  
 Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 Oy 1 PCVFIKRVNVIHG 15  
 Db 251 PSSFMRSANIVIHG 265  
 RESULT 7  
 ID SFSA\_PYRFU STANDARD; PRT; 230 AA.  
 AC Q8U1K8;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sugar fermentation stimulation protein homolog.  
 GN SFSA OR PF1198.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2261;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SFSA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE010228; AAL81322.1; -  
 DR Complete proteome.  
 KW SEQUENCE 230 AA; 26114 MW; 41B3D94D68C7A717 CRC64;  
 SQ SEQUENCE 230 AA; 26114 MW; 41B3D94D68C7A717 CRC64;  
 Query Match. 51.2%; Score 41; DB 1; Length 230;

Best Local Similarity 41.7%; Pred. No. 4.3;  
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 PCVFIKRVNVI 12  
 Db 11 PCVFIKRVNRFV 22  
 RESULT 8  
 ID LIPI\_DROME STANDARD; PRT; 439 AA.  
 AC Q46107; Q9VNR6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Lipase 1 precursor (EC 3.1.1.-).  
 GN LIPI OR CG7279.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RN SEQUENCE OF 7-433 FROM N.A.  
 RC STRAIN=Canton-S;  
 RX MEDLINE=98227315; PubMed=9566193;  
 RA Piatillo D., Marzi A., Tino A., Pilo Boyl P., Graziani F., Malva C.;  
 RT "The Drosophila melanogaster lipase homologs: a gene family with  
 RT tissue and developmental specific expression".  
 RL J. Mol. Biol. 276:877-885(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baer A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Nays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,  
 RA Foeller C., Gabrielian A.E., Garg N.S., Gilbert W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Iqbal C.,  
 RA Jaislin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Plattman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster".  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: COULD BE A DIGESTIVE ENZYME.

```

CC      EMBL; 247815; CAAB7813.1; -.
DR      SGD; S0004804; YMR192W.
DR      InterPro; IPR001195; RadGAP_TBC.
DR      Pfam; PF00566; TBC; 1.
DR      SMART; SM00164; TBC; 1.
KW      SMART; SM00164; TBC; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 720 AA; 82131 MW; E1423DB4F15F7267 CRC64;

Query Match
Best Local Similarity 48.8%; Score 39; DB 1; Length 720;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0
QY      1 PCVFIRKSVNVI 12
      |::||::|||:
Db      348 PSEYKRIISNVL 359

RESULT 10
PTPB_ECOLI STANDARD; PRT; 158 AA.
AC      P42909;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      PTS system, N-acetyl-galactosamine-specific IIB component 1 (EiIB-AGA)
DE      (N-acetyl-galactosamine-permease IIB component 1) (Phosphotransferase
DE      enzyme II, B component 1) (EC 2.7.1.69).
GN      AGAB OR B3138.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C;
RC      MEDLINE=20392444; PubMed=10931310;
RA      Birkhoetter A., Kloess H., Alpert C.-A., Lengeler J.W.;
RT      "Pathways for the utilization of N-acetyl-galactosamine and
RT      galactosamine in Escherichia coli.";
RL      Mol. Microbiol. 37:125-135(2000).
[2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;
RC      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Coliado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
[3]
RP      DISCUSSION OF SEQUENCE.
RX      MEDLINE=97086503; PubMed=8932697;
RA      Reizer J., Rameisler T.M., Reizer A., Charbit A., Sailer M.H. Jr.;
RT      "Novel phosphotransferase genes revealed by bacterial genome
RT      sequencing: a gene cluster encoding a putative N-acetyl-galactosamine
RT      metabolic pathway in Escherichia coli.";
RL      Microbiology 142:231-250(1996).
-1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT
SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
AND THE TRANSMEMBRANE CHANNEL, THE IIA DOMAIN CONTAINS THE PRIMARY
PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
THE SUGAR.
-1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
histidine + sugar phosphate
-1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
-1- SIMILARITY: CONTAINS 1 PTS EIIB DOMAIN.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```





Query Match 47.5%; Score 38; DB 1; Length 587;  
Best Local Similarity 58.3%; Pred. No. 37;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 FIKRVSNIHNG 15  
DB 327 FIKRVSNIHNG 338

## RESULT 13

POLX TOBAC STANDARD; PRT; 1328 AA.  
AC P10978;  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Retrovirus-related Pol polyprotein from transposon TMT 1-94 [Contains: Protease (EC 3.4.23.-); Reverse transcriptase (EC 2.7.7.49); Endonuclease].  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eustersids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89097311; PubMed=2536143;  
RA Gradshteyn M.-A., Spielmann A., Caboche M.;  
RT "Yn1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics."  
RL Nature 337:376-380 (1989).  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + [DNA] (N).  
CC -1- SIMILARITY: HIGH, WITH DROSOPHILA COPA ELEMENT.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.  
CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X13777; CA32025.1; -  
DR PIR: S04273; S04273.  
DR MEROPS: A11.002; -  
DR InterPro: IPR001995; Asprotease\_rtrv.  
DR InterPro: IPR001584; Rve.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam: PF00077; rvp.1.  
DR Pfam: PF00098; zf-CCHC; 1.  
DR Pfam: PF00685; tve.1.  
DR PRINTS: PR00939; C2HCZNFINGER.  
DR SMART: SM00343; Znf\_CCHC; 1.  
DR PROSITE: PS50158; ZF\_CCHC; 1.  
KW Hydroxylase; Aspartyl protease; RNA-directed DNA polymerase; Endonuclease; Transferase; Polyprotein; Transposable element;  
KW Zinc-finger.  
FT ZN FING 230 247 CCHC-TYPE.  
FT ACT SITE 297 297 PROTEASE (BY SIMILARITY).  
SQ SEQUENCE 1328 AA; 151076 MW; F27E76C504B19B1B CRC64;

Query Match 47.5%; Score 38; DB 1; Length 1328;  
Best Local Similarity 60.0%; Pred. No. 81;  
Matches 9; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 1 PCVFIKVS--NVII 13  
DB 990 PCVFIKVSNNFII 1004

RESULT 14  
SP22\_BACSH STANDARD; PRT; 146 AA.  
ID SP22\_BACSH  
AC O32724;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-1998 (Rel. 36, Last annotation update)  
DE Anti-sigma F factor (Stage II sporulation protein AB).  
GN SPOIIB.  
OS Bacillus sphaericus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1421;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=2362;  
RX MEDLINE=97409954; PubMed=9266669;  
RA Park S.G., Yuckin M.D.;  
RT "Sequencing and phylogenetic analysis of the spoIIA operon from diverse Bacillus and Paenibacillus species."  
RL Gene 194:25-33 (1997).  
CC -1- FUNCTION: BINDS TO SIGMA F AND BLOCKS ITS ABILITY TO FORM AN RNA POLYMERASE HOLOENZYME (E-SIGMA F). PHOSPHORYLATES SPOIIA ON A SERINE RESIDUE. THIS PHOSPHORYLATION MAY ENABLE SPOIIA TO ACT AS AN ANTI-ANTI-SIGMA FACTOR THAT COUNTERACTS SPOIIB AND THUS RELEASES SIGMA F FROM INHIBITION (BY SIMILARITY).  
CC -1- SIMILARITY: STRONG, TO RBM.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: L4759; AAB8190.1; -  
DR InterPro: IPR003594; ATPbind\_ATPase.  
DR InterPro: IPR004359; HIS\_KIN\_sig.  
DR Pfam: PF02518; HATPase\_c1.  
DR SMART: SM00387; HATPase\_c1.  
KW Sporulation; Transferase; Kinase.  
SQ SEQUENCE 146 AA; 16124 MW; FB56849D67C20CAC CRC64;

Query Match 46.2%; Score 37; DB 1; Length 146;  
Best Local Similarity 61.5%; Pred. No. 14;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 VFIRVSNVIHNG 15  
DB 43 VVSEAVSNVIHNG 55

## RESULT 15

ATNC\_HUMAN STANDARD; PRT; 290 AA.  
AC P14415; O60444;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Sodium/potassium-transferring ATPase beta-2 chain (sodium/potassium-dependent ATPase beta-2 subunit).  
GN ATP1B2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89174720; PubMed=2538450;  
RA Martin-Vasallo P., Dackowski W., Emanuel J.R., Levenson R.;  
RT "Identification of a putative isoform of the Na,K-ATPase beta subunit. Primary structure and tissue-specific expression."  
RL J. Biol. Chem. 264:4613-4618 (1989).



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds  
(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524D-36  
Perfect score: 80  
Sequence: 1 PCVFIRKRVSNVTHG 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*

2: SP\_BACTERIA:\*

3: SP\_FUNGI:\*

4: SP\_HUMAN:\*

5: SP\_INVERTEBRATE:\*

6: SP\_MAMMAL:\*

7: SP\_MHC:\*

8: SP\_ORGANELLE:\*

9: SP\_PHAGE:\*

10: SP\_PLANT:\*

11: SP RODENT:\*

12: SP\_VIRUS:\*

13: SP\_VERTEBRATE:\*

14: SP\_UNCLASSIFIED:\*

15: SP\_VIRUS:\*

16: SP\_BACTERIA:\*

17: SP\_ARCHAEA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	374	10	Q8RUR1
2	61	76.2	367	10	Q9LUT2
3	61	76.2	367	10	Q9LUT2
4	61	76.2	367	10	Q9LUT1
5	61	76.2	367	10	Q9M4S6
6	61	76.2	367	10	Q9M4S5
7	61	76.2	367	10	Q9M4S4
8	61	76.2	367	10	Q9M4S3
9	61	76.2	367	10	Q9M4S2
10	57	71.2	367	10	Q93X51
11	56	70.0	390	10	Q65388
12	56	70.0	326	10	Q23667
13	56	70.0	394	10	Q88345
14	56	70.0	394	10	Q65457
15	56	70.0	398	10	Q65456
16	56	70.0	404	10	Q43783
					Q9SB71

17	56	70.0	407	10	Q9SDW4	Q9SDW4 musa acumin
18	56	70.0	408	10	Q9C5M8	Q9C5M8 arabidopsis
19	56	70.0	409	10	Q93XJ1	Q93XJ1 salix gilgi
20	56	70.0	410	10	Q9FMK5	Q9FMK5 arabidopsis
21	56	70.0	432	10	Q93Z25	Q93Z25 arabidopsis
22	54	67.5	367	10	Q93XR6	Q93XR6 cupressus a
23	53	66.2	398	10	Q9M505	Q9M505 vitis vinif
24	53	66.2	401	10	Q24554	Q24554 zinnia eleg
25	52	65.0	368	10	Q9C8G4	Q9C8G4 arabidopsis
26	50	62.5	452	10	Q9LRM5	Q9LRM5 arabidopsis
27	49	61.3	405	10	Q24416	Q24416 fragaria an
28	49	61.3	405	10	Q94FR6	Q94FR6 fragaria an
29	49	61.3	449	10	Q91J42	Q91J42 arabidopsis
30	48	61.3	454	10	Q9SDW3	Q9SDW3 musa acumin
31	48	60.0	143	10	Q94B14	Q94B14 vitis vinif
32	48	60.0	181	10	Q23668	Q23668 arabidopsis
33	48	60.0	227	10	Q23668	Q23668 arabidopsis
34	48	60.0	354	10	Q94001	Q94001 arabidopsis
35	48	60.0	368	10	Q94FT5	Q94FT5 arabidopsis
36	48	60.0	374	10	Q9SVF1	Q9SVF1 fragaria an
37	48	60.0	392	10	Q9FXD8	Q9FXD8 arabidopsis
38	48	60.0	393	10	Q9F181	Q9F181 arabidopsis
39	48	60.0	408	10	Q8W116	Q8W116 arabidopsis
40	48	60.0	417	10	Q93WF1	Q93WF1 arabidopsis
41	48	60.0	431	10	Q23017	Q23017 arabidopsis
42	48	60.0	470	10	Q944R1	Q944R1 arabidopsis
43	47	58.8	409	10	Q9L720	Q9L720 arabidopsis
44	46	57.5	416	10	Q9M828	Q9M828 arabidopsis
45	46	57.5	418	10	Q9SVQ6	Q9SVQ6 arabidopsis

## ALIGNMENTS

RESULT 1	ID	Q8RUR1	PRELIMINARY;	PRT;	374 AA.
AC	Q8RUR1	01-JUN-2002 (T-EMBLrel. 21, Created)			
DT	01-JUN-2002 (T-EMBLrel. 21, Last sequence update)				
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)				
DE	Cry j 1 precursor.				
GN	Cry j 1.1 OR Cry j 1.2.				
OS	Cryptomeria japonica (Japanese cedar).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Coniferales; Cupressaceae; Cryptomeria.				
OX	NCBI_TaxID=3369;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=POLLEN;				
RA	Futamura N., Shinohara K.;				
RT	"Isolation and characterization of cDNAs encoding major allergen Cry j				
RT	1 from Cryptomeria japonica pollen."				
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB081309; BAB86286.1; -				
DR	EMBL; AB081310; BAB86287.1; -				
KW	Signal.				
FT	SIGNAL	1	21	POTENTIAL.	
FT	CHAIN	22	374	Cry j 1.	
SQ	SEQUENCE	374 AA;	40720 MW;	90D0085D24BF2BD4 CRC64;	
Query Match					
Best Local Similarity 100.0%; Score 80; DB 10; Length 374;					
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	PCVFIRKRVSNVTHG 15			
DB	127	PCVFIRKRVSNVTHG 141			
RESULT 2					
Q9LUT2					
ID	Q9LUT2	PRELIMINARY;	PRT;	367 AA.	

AC 09LRT2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-2.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21315424; PubMed=11422137;  
 RA Mido-Horiuti T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Exp. Allergy 31:771-778 (2001).  
 DR EMBL; AF151427; AAF80164.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39768 MW; 0A6AC2F1BAF89586 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVILHG 15  
 DB 127 PCLFMKRVSHVILHG 141

RESULT 3  
 09LRT1 PRELIMINARY; PRT; 367 AA.

ID 09LRT1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1-1.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21315424; PubMed=11422137;  
 RA Mido-Horiuti T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Exp. Allergy 31:771-778 (2001).  
 DR EMBL; AF151427; AAF80164.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39708 MW; DCBD1981A74E4711 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVILHG 15  
 DB 127 PCLFMKRVSHVILHG 141

RESULT 4  
 09M4S6 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.

GN CUP81.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257491; AAF72625.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39835 MW; B4B9C60108C2C5A3 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVILHG 15  
 DB 127 PCLFMKRVSHVILHG 141

RESULT 5  
 09M4S5 PRELIMINARY; PRT; 367 AA.

ID 09M4S5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP81.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257492; AAF72626.1; -  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PRO0807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39894 MW; SD56FC03263B741 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVILHG 15  
 DB 127 PCLFMKRVSHVILHG 141

RESULT 6  
 09M4S4 PRELIMINARY; PRT; 367 AA.  
 AC 09M4S4;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP81.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 NCBI\_TaxID=13469;

[1]  
 RN Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RP EMBL, AF257495; AAF72629.1; -  
 RA InterPro; IPR002022; Amb allergen.  
 RT Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR SEQUENCE 367 AA; 39934 MW; 974D3011D74E3D6E CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVITIG 15  
 DB 127 PCLFMKRVSHVILHG 141

## RESULT 7

OQ9M4S3 PRELIMINARY; PRT; 367 AA.  
 AC OQ9M4S3.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSL.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR SEQUENCE 367 AA; 39832 MW; BSDPFB5A61C07A53 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVITIG 15  
 DB 127 PCLFMKRVSHVILHG 141

## RESULT 8

OQ9M4S2 PRELIMINARY; PRT; 367 AA.  
 AC OQ9M4S2.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUPSL.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";

RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF257495; AAF72629.1; -  
 RA InterPro; IPR002022; Amb allergen.  
 RT Pfam; PF00544; pec lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR SEQUENCE 367 AA; 39819 MW; AE7E055A61C07A53 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVITIG 15  
 DB 127 PCLFMKRVSHVILHG 141

## RESULT 9

OQ93X51 PRELIMINARY; PRT; 367 AA.  
 AC OQ93X51.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Putative allergen jun o 1.  
 GN JUN O 1.  
 OS Juniperus oxycedrus (Prickly juniper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxID=99008;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLLEN;  
 RA Taciavacci P., Di Felice G., Pini C.;  
 RT "Cloning of Juniperus oxycedrus major allergen.";  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ293767; CAC48400.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec lyase; 1.  
 DR SEQUENCE 367 AA; 39808 MW; 5D28204DBFD1B9D7 CRC64;

Query Match 76.2%; Score 61; DB 10; Length 367;  
 Best Local Similarity 60.0%; Pred. No. 0.006;  
 Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVITIG 15  
 DB 127 PCLFMKRVSHVILHG 141

## RESULT 10

O65388 PRELIMINARY; PRT; 390 AA.  
 AC O65388.  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE F12P1.22 protein.  
 GN F12P1.22.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Vysotskaia V.S., Osborne B.I., Schwartz J.R., Tortum M., Yu G.,  
 RA Kwan A., Oji O., Liu S., Buehler E., Conway A.B., Conway A.R.,  
 RA Dewar K., Feng J., Kim C., Kurtz D., Palm C.J., Li Y., Shinn P.,  
 RA Sun H., Davis R.W., Ecker J.R., Federpfiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F12P1 sequence, complete  
 sequence";  
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC002131; AAC17625.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 390 AA; 43354 MW; 7E4D3CA9BB396CB7 CRC64;

Query Match 71.2%; Score 57; DB 10; Length 390;  
 Best Local Similarity 66.7%; Pred. No. 0.035;  
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 PCVFIKRVSNVTHG 15  
 DB 140 PCITIQVTVNIHIG 154

## RESULT 11

ID 023667 PRELIMINARY; PRT; 226 AA.  
 AC 023667;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Putative pectate lyase (Fragment).  
 GN A10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97422403; PubMed=9278171;  
 RA Kulkarni S., McCormick S.;  
 RT "Identification of the tobacco and Arabidopsis homologues of the  
 RL pollen-expressed LATE5 gene of tomato.";  
 RL Plant Mol Biol. 34:809-814(1997).  
 DR EMBL; U83621; AAB69761.1; -  
 DR HSSP; P39116; 1BN8.  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Lyase.  
 FT NON\_TER 1  
 SQ SEQUENCE 226 AA; 25103 MW; F999B38285D645C CRC64;

Query Match 70.0%; Score 56; DB 10; Length 226;  
 Best Local Similarity 60.0%; Pred. No. 0.031;  
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PCVFIKRVSNVTHG 15  
 DB 92 PCITIQVTVNIHIG 106

## RESULT 12

ID 083345 PRELIMINARY; PRT; 324 AA.  
 AC 083345;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative pectate lyase (Fragment).  
 GN UP48.  
 OS Capsicum annuum (Bell pepper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.  
 OC NCBI\_TaxID=4072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Marois E., van den Ackerveken G., Binas U.;  
 RT "The Xanthomonas type III effector protein AvrBs3 modulates plant gene

RT expression and induces cell hypertrophy in the susceptible host.";  
 RL Mol. Plant Microbe Interact. 0:0-0(2002).  
 DR EMBL; AF492632; AAM12784.1; -  
 KW Lyase.  
 FT NON\_TER 324  
 SQ SEQUENCE 324 AA; 35751 MW; BC2BED36FB96B7A CRC64;

Query Match 70.0%; Score 56; DB 10; Length 324;  
 Best Local Similarity 60.0%; Pred. No. 0.044;  
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PCVFIKRVSNVTHG 15  
 DB 158 PCITIQVTVNIHIG 172

## RESULT 13

ID 065457 PRELIMINARY; PRT; 394 AA.  
 AC 065457;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Pectate lyase like protein.  
 GN FIN20.190 OR AT4G22090.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Medler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,  
 RA Schueller C., Wambutt R. to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAR-1998)  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Medler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL022140; CAA18112.1; -  
 DR EMBL; AL161556; CAB79164.1; -  
 DR InterPro; IPR002022; Amb allergen.  
 DR Pfam; PF00544; pec\_lyase1.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 394 AA; 43299 MW; 3912C1F6A8C6E3F5 CRC64;

Query Match 70.0%; Score 56; DB 10; Length 394;  
 Best Local Similarity 60.0%; Pred. No. 0.054;  
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PCVFIKRVSNVTHG 15  
 DB 150 PCITIQVTVNIHIG 164

## RESULT 14

ID 065456 PRELIMINARY; PRT; 394 AA.  
 AC 065456;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Pectate lyase like protein.  
 GN FIN20.180 OR AT4G22080.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCB1\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,

RA Scheller C.; Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.,

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL022140; CA1811.1; -

DR EMBL; AL161556; CAB79163.1; -

DR InterPro; IPR002022; Amb allergen.

DR InterPro; IPR002048; EF-hand.

DR Pfam; PF00544; pec\_lyase; 1.

DR PRINTS; PRO0807; AMBALERGEN.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.

SO SEQUENCE 394 AA; 43476 MW; A48BD586ECF148CA CRC64;

Query Match

Best Local Similarity 70.0%; Score 56; DB 10; Length 394;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 PCVFIKRVSNVING 15

DB 150 PCITIQDVNTNVIHG 164

RESULT 15

ID Q43783

AC Q43783; PRELIMINARY; PRT; 398 AA.

DT 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

PE Pectate lyase (EC 4.2.2.2) (Fragment).

GN PEL.

OS Musa acuminata (Banana).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;

OC Musa.

OX NCB1\_TaxID=4641;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=CV. DWARF CAVENDISH; TISSUE=RIPE FRUIT;

RA Dominguez-Piñan E.; Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA STRAIN=CV. DWARF CAVENDISH; TISSUE=RIPE FRUIT;

RA Dominguez-Piñan E.; Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; X92943; CA63496.1; -

DR HSP; P39116; I8N8.

DR InterPro; IPR002022; Amb allergen.

DR Pfam; PF00544; pec\_lyase; 1.

DR PRINTS; PRO0807; AMBALERGEN.

DR Lyase.

FT NON TER

SO SEQUENCE 398 AA; 43712 MW; A830293E28C8362C CRC64;

Query Match

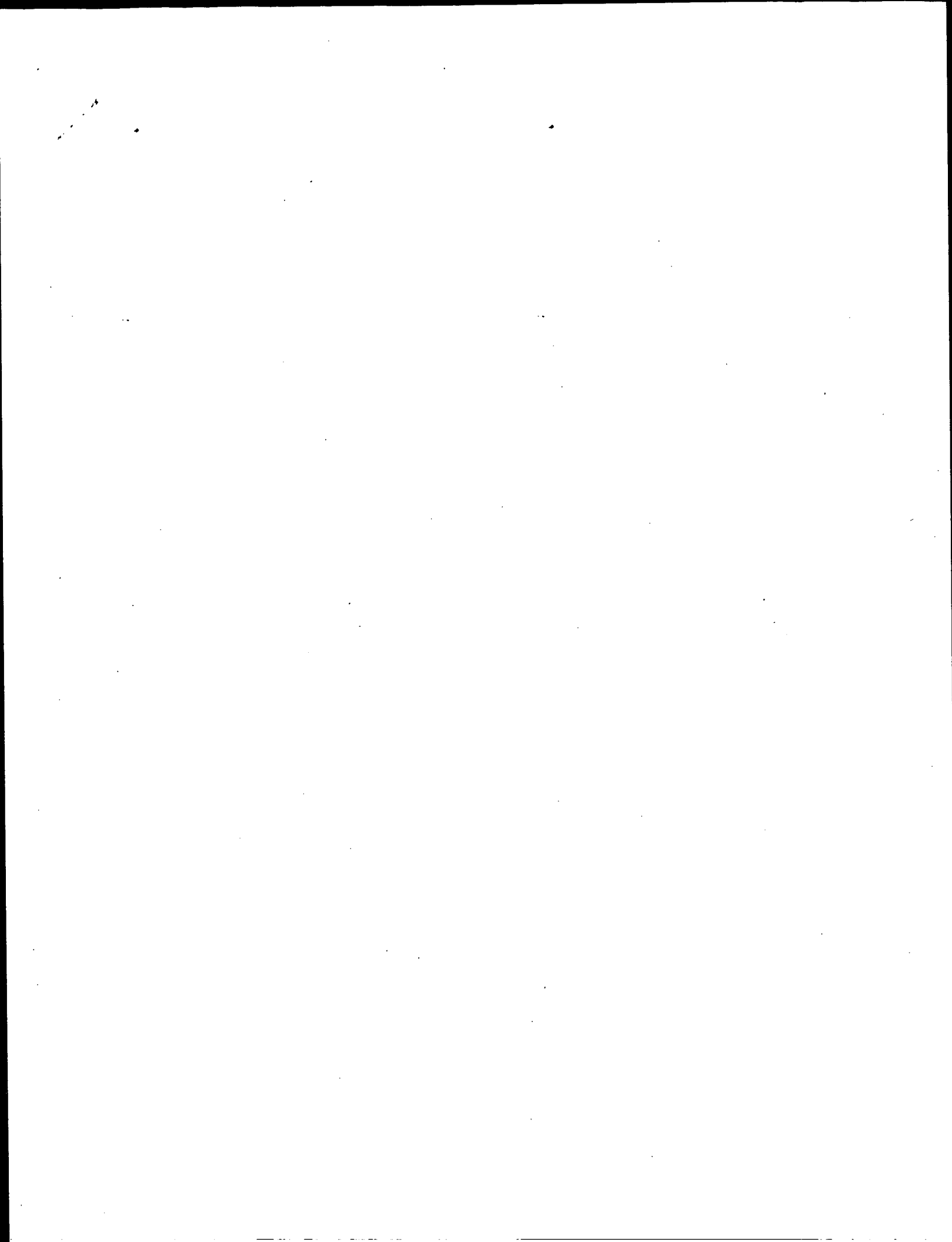
Best Local Similarity 70.0%; Score 56; DB 10; Length 398;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 PCVFIKRVSNVING 15

DB 153 PCITIQDVNTNVIHG 167

Search completed: April 20, 2003, 13:13:06  
Job time : 13.6711 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524D-57  
Perfect score: 78  
Sequence: 1 KSMKVTVAENQFGPN 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	78	100.0	374 2 UC2124	major allergen Cry
2	78	100.0	374 2 UC2123	major allergen Cry
3	58	74.4	368 2 G86427	probable pectate 1
4	58	74.4	394 2 T49115	pectate lyase like
5	58	74.4	394 2 T49116	pectate lyase like
6	57	73.1	390 2 H86253	hypothetical prote
7	52	66.7	404 2 S32209	pectate lyase (EC
8	52	66.7	404 2 T05536	pectate lyase (EC
9	51	65.4	455 2 T00856	pectate lyase (EC
10	51	65.4	459 2 G86278	pectate lyase (EC
11	49	62.8	397 2 S26211	hypothetical prote
12	49	62.8	431 2 F86179	hypothetical prote
13	49	62.8	431 2 T09524	probable pectate 1
14	48	61.5	472 2 T11456	pectate lyase-like
15	47	60.3	418 2 T07701	pectate lyase (EC
16	47	60.3	434 2 S29612	pectate lyase (EC
17	47	60.3	438 2 S43335	pectate lyase (EC
18	46	59.0	374 2 T05240	pectate lyase (EC
19	46	59.0	374 2 H85148	pectate lyase (EC
20	46	59.0	398 2 T07058	probable pectate 1
21	46	59.0	463 2 T46165	pectate lyase (EC
22	46	59.0	542 2 T06728	pectate lyase (EC
23	43	55.1	449 2 S27058	pectate lyase (EC
24	42	53.8	397 2 C53240	pectate lyase (EC
25	42	53.8	397 2 C39039	allergen Amb a 1.3
26	42	53.8	1199 2 S77082	pyruvate (flavodox
27	41	52.6	416 2 S45484	glycerinaldehyde-3-p
28	41	52.6	2823 2 T23064	hypothetical prote
29	41	52.6	2823 2 F87908	protein T22A3.8 [1

30	41	52.6	3102 2	T43291
31	40	51.3	333 2	I84743
32	40	51.3	336 2	I49766
33	40	51.3	363 2	G95237
34	40	51.3	363 2	H98101
35	40	51.3	677 2	B71870
36	39	50.0	210 2	D66531
37	39	50.0	227 2	B90429
38	39	50.0	286 2	T09704
39	39	50.0	348 2	S77104
40	39	50.0	392 2	D53240
41	39	50.0	398 2	B53240
42	39	50.0	398 2	B53240
43	38	48.7	329 2	B39099
44	38	48.7	329 2	G82785
45	38	48.7	364 2	AH1857
			396 2	A39099

## ALIGNMENTS

## RESULT 1

UC2124 major allergen Cry j I precursor (clone PCCI-15) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: UC2124

R/Source, T: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan

A/Reference number: UC2123; MUID:94183234; PMID:8155802

A/Accession: UC2124

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:026545; NID:9493633; PIDN:BAA05543.1; PID:9493634

A/Experimental source: pollen

A/Note: the authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase LRT59

C/Keywords: glycoprotein; pollen

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-374/Product: major allergen Cry j I (clone PCCI-15) #status predicted <MAT>

F/158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 78; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. No. 8.4e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAENQFGPN 15

DB 232 KSMKVTVAENQFGPN 246

## RESULT 2

UC2123 major allergen Cry j I precursor (clone PCCI-2-2) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C/Accession: UC2123; PC2065

R/Source, T: Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 199, 619-625, 1994

A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japan

A/Reference number: UC2123; MUID:94183234; PMID:8155802

A/Accession: UC2123

A/Molecule type: mRNA

A/Residues: 1-374 <SON>

A/Cross-references: GB:026544; NID:9493631; PIDN:BAA05542.1; PID:9493632

A/Experimental source: pollen

A/Accession: PC2065

A/Molecule type: protein

A/Residues: 22-53;58-81;219-232;236-258;299-307;346-372 <SO>

A/Note: the authors described carbohydrate binding site for residue 279

C/Superfamily: pectate lyase LRT59

C/Keywords: glycoprotein; pollen

C/Keywords: glycoprotein; pollen

F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen cary j i (clone pCCT-2-2) #status predicted <MAT>  
F:158,191,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 78; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 8,4e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
DB 232 KSMKVTVAFNQFGPN 246

## RESULT 3

probable pectate lyase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: G86427

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86427

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-368 <STO>

A:Cross-references: GB:AE005172; NID:g11055821; PIDN:AA628291.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: pectate lyase LAT59

Query Match 74.4%; Score 58; DB 2; Length 368;  
Best Local Similarity 71.4%; Pred. No. 0.0051;

Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGP 14  
DB 229 KSMKVTVAFNQFGP 242

## RESULT 4

pectate lyase like protein - Arabidopsis thaliana

N:Alternate names: protein AT4g22080

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Sep-2000

C:Accession: T49115

R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; M

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25016

A:Accession: T49115

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <BEV>

A:Cross-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4g22080

A:Experimental source: cultivar Columbia; BAC clone FIN20

C:Genetics:

A:Gene: ATSP:AT4g22080

A:Map position: 4

A:Introns: 240/3; 275/2; 321/3

C:Superfamily: pectate lyase LAT59

## Query Match

Best Local Similarity 74.4%; Score 58; DB 2; Length 394;  
Best Local Similarity 91.7%; Pred. No. 0.0055;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 MKVTVAFNQFGP 14  
DB 256 MKVTVAFNQFGP 267

## RESULT 5

pectate lyase like protein - Arabidopsis thaliana

N:Alternate names: protein AT4g22090

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000

C:Accession: T49116

R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.;

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25016

A:Accession: T49116

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <BEV>

A:Cross-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4g22090

A:Experimental source: cultivar Columbia; BAC clone FIN20

C:Genetics:

A:Gene: ATSP:AT4g22090

A:Map position: 4

A:Introns: 240/3; 275/2; 321/3

C:Superfamily: pectate lyase LAT59

Query Match 74.4%; Score 58; DB 2; Length 394;  
Best Local Similarity 91.7%; Pred. No. 0.0055;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 MKVTVAFNQFGP 14  
DB 256 MKVTVAFNQFGP 267

## RESULT 6

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: H86253

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H86253

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-390 <STO>

A:Cross-references: GB:AE005172; NID:g157942; PIDN:AA17625.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: pectate lyase LAT59

Query Match 73.1%; Score 57; DB 2; Length 390;  
Best Local Similarity 71.4%; Pred. No. 0.0084;

Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGP 14  
DB 244 KSMKVTVAFNQFGP 257

## RESULT 7

S12209  
pectate lyase (EC 4.2.2.2) - tomato  
C/Species: Lycopersicon esculentum (tomato)  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999  
C/Accession: S12209  
R/Budeller, K.A.; Smith, A.G.; Gasser, C.S.  
Mol. Gen. Genet. 224, 183-192, 1990  
A/Title: Regulation of a stylar transmitting tissue-specific gene in wild-type and trans  
A/Reference number: S12209; MUID:91117185; PMID:227637  
A/Accession: S12209  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-404 <BUD>  
A/Cross-references: GB:X55193; NID:G19161; PIDN:CAA38979.1; PID:G19162  
C/Keywords: carbon-oxygen lyase

Query Match 66.7%; Score 52; DB 2; Length 404;  
Best Local Similarity 76.9%; Pred. No. 0.077;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
|:|:|:|:|:|:|  
Db 262 KMQVTVAFNHFG 274

## RESULT 8

T05556  
pectate lyase (EC 4.2.2.2) P22K18.20 - Arabidopsis thaliana  
N/Alternate names: protein P22K18.20  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C/Accession: T05556  
R/Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohelsel, J.; Mewes, H.W.; Meyer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A/Reference number: Z15419  
A/Accession: T05556  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-404 <BEV>  
A/Cross-references: EMBL:AL035356  
A/Experimental source: cultivar Columbia; BAC clone P22K18  
C/Genetics:  
A/Map position: 4  
A/Intons: 37/2; 287/2  
A/Note: P22K18.20  
C/Superfamily: pectate lyase LAT59  
C/Keywords: carbon-oxygen lyase

Query Match 66.7%; Score 52; DB 2; Length 404;  
Best Local Similarity 69.2%; Pred. No. 0.077;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
|:|:|:|:|:|:|  
Db 266 KMQVTVAFNHFG 278

## RESULT 9

T00856  
pectate lyase (EC 4.2.2.2) T20F6.14 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Feb-2001  
C/Accession: T00856; AB4440  
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, March 1998  
A/Description: Arabidopsis thaliana chromosome II BAC T20F6 genomic sequence.  
A/Reference number: Z14206  
A/Accession: T00856  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-455 <ROU>  
A/Cross-references: EMBL:AC002521; NID:G2947056; PIDN:AAC05350.1; PID:G2947069  
A/Experimental source: cultivar Columbia

R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, T.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: AB4440; MUID:20083487; PMID:10617197  
A/Accession: AB4440  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-455 <STO>  
A/Cross-references: GB:AE002093; NID:G2947069; PIDN:AAC05350.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: AC2902720  
A/Map position: 2  
A/Intons: 66/2; 295/3; 376/3  
A/Note: T20F6.14  
C/Superfamily: pectate lyase LAT59  
C/Keywords: carbon-oxygen lyase

Query Match 65.4%; Score 51; DB 2; Length 455;  
Best Local Similarity 69.2%; Pred. No. 0.14;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
|:|:|:|:|:|:|  
Db 309 KMQVTVAFNHFG 321

## RESULT 10

G86278  
hypothetical protein F14L17.19 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 01-Mar-2002  
C/Accession: G86278  
R/Theologis, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alom  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, B.  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: AB6141; MUID:21016719; PMID:11130712  
A/Accession: G86278  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-459 <STO>  
A/Cross-references: GB:AE005172; NID:G7262684; PIDN:AAF43942.1; GSPDB:GN00141  
C/Genetics:  
A/Map position: 1  
C/Superfamily: pectate lyase LAT59

Query Match 65.4%; Score 51; DB 2; Length 459;  
Best Local Similarity 69.2%; Pred. No. 0.14;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
|:|:|:|:|:|:|  
Db 313 KMQVTVAFNHFG 325

## RESULT 11

G26211  
pectate lyase (EC 4.2.2.2) - common tobacco  
C/Species: Nicotiana tabacum (common tobacco)  
C/Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 21-Jul-2000  
C/Accession: S26211; S26212; S21933; S22753; S22754  
R/Rogers, H.J.; Harvey, A.; Lonsdale, D.M.  
Plant Mol. Biol. 20, 493-502, 1992  
A/Title: Isolation and characterization of a tobacco gene with homology to pectate lyase  
A/Reference number: S26211; MUID:93043039; PMID:1421152

A:Accession: S26211  
 A:Molecule type: DNA  
 A:Residues: 1-397 <ROG>  
 A:Cross-references: EMBL:X67158; NID:g19907; PIRN:CAA47630.1; PID:g19908  
 A:Accession: S26212  
 A:Molecule type: mRNA  
 A:Residues: 119-155, 'C', 157-188, 190, 'G', 191-199, 'D', 200, 'R', 203-248, 'N', 250-381 <RO2>  
 A:Cross-references: EMBL:X67159; NID:g19909; PIRN:CAA47631.1; PID:g3980174  
 A:Note: translation of the nucleotide sequence is not complete  
 R:Lonsdale, D.M.  
 Submitted to the EMBL Data Library, July 1991  
 A:Reference number: S21933  
 A:Accession: S21933  
 A:Molecule type: DNA  
 A:Residues: 1-397 <LON>  
 A:Cross-references: EMBL:X61102; NID:g19981; PIRN:CAA43414.1; PID:g19982  
 C:Genetics:  
 A:Introns: 193/1, 293/2  
 C:Superfamily: pectate lyase LAT59  
 C:Keywords: carbon-oxygen lyase

Query Match 62.8%; Score 49; DB 2; Length 397;  
 Best Local Similarity 61.5%; Pred. No. 0.28;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
 ||:|||||  
 Db 254 KMQVTVAFNHFG 266

## RESULT 12

hypothetical protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: F86179  
 R:Neolagus, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hulst, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maftl, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F86179  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-431 <STO>  
 A:Cross-references: GB:AEO05172; NID:g2494113; PIRN:AAB80622.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: pectate lyase LAT59

Query Match 62.8%; Score 49; DB 2; Length 431;  
 Best Local Similarity 61.5%; Pred. No. 0.31;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
 ||:|||||  
 Db 289 KMQVTVAFNHFG 301

## RESULT 13

T09524  
 probable pectate lyase (EC 4.2.2.2) - alfalfa  
 C:Species: Medicago sativa (alfalfa)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 26-Aug-1999  
 C:Accession: T09524  
 R:Wu, Y.; Qiu, X.; Du, S.; Erickson, L.  
 submitted to the EMBL Data Library, November 1995

A:Reference number: Z16712  
 A:Accession: T09524  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-450 <MOV>  
 A:Cross-references: EMBL:U41472; NID:g1171160; PID:g1171161  
 C:Genetics:  
 A:Introns: 65/2, 295/3  
 C:Function:  
 A:Description: catalyzes cleavage of pectate to oligosaccharides  
 C:Superfamily: pectate lyase LAT59  
 C:Keywords: carbon-oxygen lyase

Query Match 62.8%; Score 49; DB 2; Length 450;  
 Best Local Similarity 60.2%; Pred. No. 0.32;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
 ||:|||||  
 Db 309 KMQVTVAFNHFG 321

## RESULT 14

T51456  
 pectate lyase-like protein - Arabidopsis thaliana  
 N:Alternate names: protein F2G14.230  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 17-Nov-2000  
 C:Accession: T51456  
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; N  
 submitted to the Protein Sequence Database, August 2000  
 A:Reference number: Z25394  
 A:Accession: T51456  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-472 <SAT>  
 A:Cross-references: EMBL:AL391146  
 A:Experimental source: cultivar Columbia; BAC clone F2G14  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 65/2; 316/3; 398/3  
 A:Note: F2G14.230  
 C:Superfamily: pectate lyase LAT59

Query Match 61.5%; Score 48; DB 2; Length 472;  
 Best Local Similarity 60.2%; Pred. No. 0.52;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFG 13  
 ||:|||||  
 Db 330 KMQVTVAFNHFG 342

## RESULT 15

T07701  
 pectate lyase (EC 4.2.2.2) F17N18.100 - Arabidopsis thaliana  
 N:Alternate names: protein F17N18.100  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 23-Jul-1999  
 C:Accession: T07701  
 R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Meyer, K.F.X.; Sch  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16098  
 A:Accession: T07701  
 A:Molecule type: DNA  
 A:Residues: 1-418 <BEV>  
 A:Cross-references: EMBL:AL049751  
 A:Experimental source: cultivar Columbia; BAC clone F17N18  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 26/2; 49/2; 264/3; 299/2; 345/3  
 A:Note: F17N18.100  
 C:Superfamily: pectate lyase LAT59

C/Keywords: carbon-oxygen lyase

Query Match

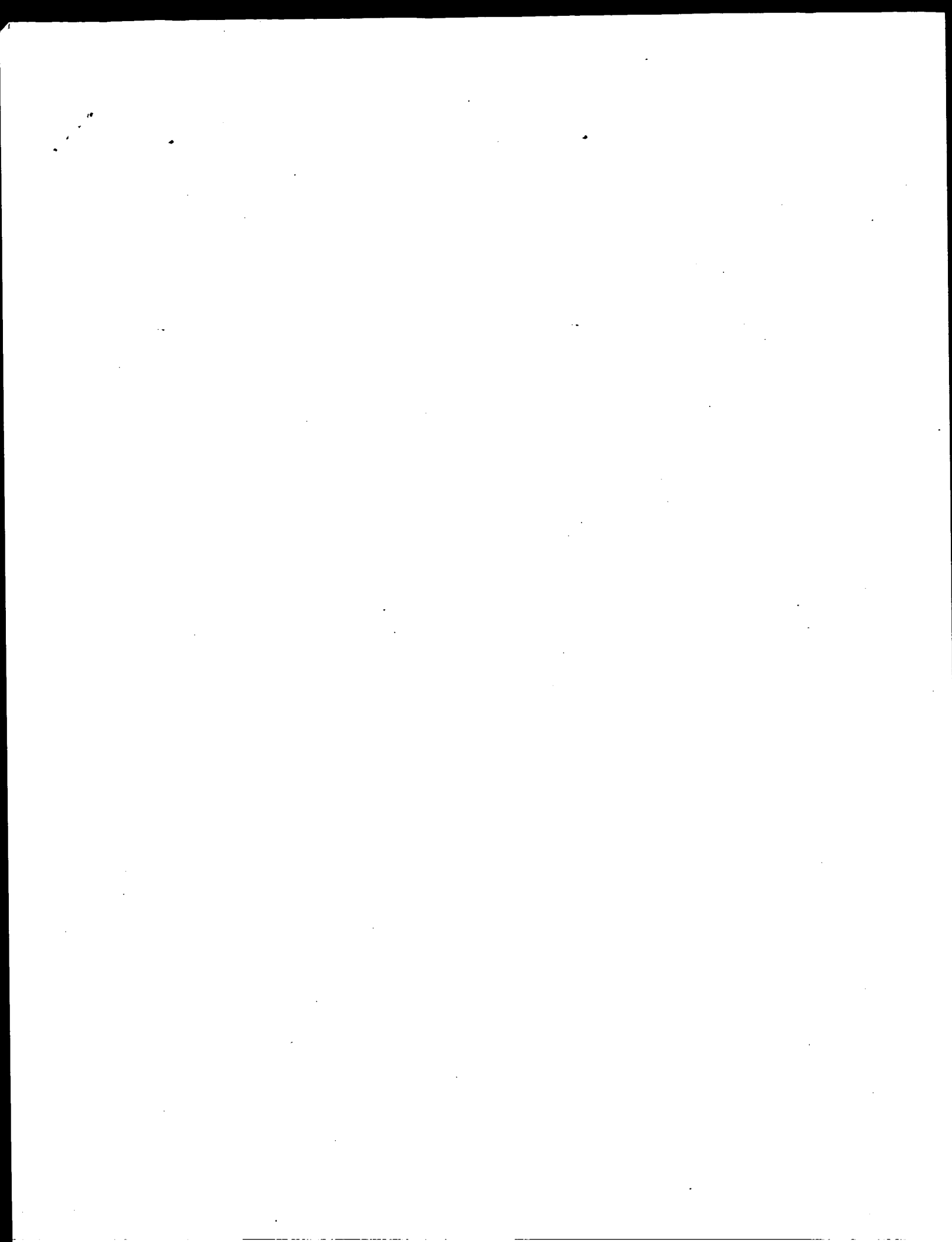
Best Local Similarity 60.3%; Score 47; DB 2; Length 418;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAENQFG 13

Db 278 KMMQVTIAYNHFG 290

Search completed: April 20, 2003, 13:15:46  
Job time : 6.07895 secs



GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524D-57

Perfect score: 78

Sequence: 1 KSMKVTVAENQFGPN 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	78	100.0	346	1	MPAL_CUPAR
2	78	100.0	367	1	MPAL_JUNAS
3	78	100.0	374	1	SBR_CRYDA
4	78	100.0	375	1	MPAL_CHAOB
5	52	66.7	404	1	PEL12_LYCES
6	49	62.8	397	1	PEL12_LYCES
7	47	60.3	434	1	PEL12_LYCES
8	46	59.0	398	1	PEL12_LYCES
9	43	55.1	449	1	PEL12_LYCES
10	42	53.8	397	1	MP13_AMBAR
11	42	53.8	1199	1	MP13_AMBAR
12	41	52.6	416	1	G3PA_GRAVE
13	40	51.3	333	1	EBF2_HUMAN
14	40	51.3	336	1	EBF2_MOUSE
15	39	50.0	327	1	EBF1_XENLA
16	39	50.0	348	1	TRPD_SYNY3
17	39	50.0	392	1	MP14_AMBAR
18	39	50.0	398	1	MP12_AMBAR
19	39	50.0	416	1	O35A_DROME
20	38	48.7	332	1	EBF2_BRARE
21	38	48.7	396	1	MP11_AMBAR
22	38	48.7	414	1	G3PA_CHOCR
23	38	48.7	466	1	YAM8_SCHRO
24	38	48.7	589	1	TRPD_THEMA
25	37	47.4	127	1	LYC_COLLI
26	37	47.4	272	1	TRA2_DROVI
27	37	47.4	334	1	EBF1_CHICK
28	37	47.4	340	1	G3P2_BACSU
29	37	47.4	345	1	EBF1_MOUSE
30	37	47.4	345	1	EBF1_RAT
31	37	47.4	346	1	EBF1_HUMAN
32	37	47.4	367	1	TGM2_CHICK
33	37	47.4	716	1	ZP2_FELCA

34	36.5	46.8	1126	1	V125_AMVLE
35	36.5	46.8	1248	1	TOFG_SULAC
36	36.5	46.8	3075	1	LMAL_HUMAN
37	36	46.2	258	1	LYCV_BPPH2
38	36	46.2	258	1	LYCV_BPPH2
39	36	46.2	331	1	TRPD_BUCSC
40	36	46.2	444	1	TRPD_ARATH
41	36	46.2	560	1	YDEN_ECOLI
42	36	46.2	823	1	CNBA_MOUSE
43	35	44.9	115	1	YE69_HAETN
44	35	44.9	173	1	CME_HAETN
45	35	44.9	176	1	SLYD_TREPA

## ALIGNMENTS

## RESULT 1

MPAL\_CUPAR STANDARD; PRT; 346 AA.

ID MPAL\_CUPAR STANDARD; PRT; 346 AA.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Major pollen allergen Cup a 1.

OS Cupressus arizonica.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferales; Cupressaceae; Cupressus.

ON NCBI\_Taxid=49011;

RP SEQUENCE FROM N.A.

RX MEDLINE=20571526; PubMed=1112214;

RA Aceituno E., Del Pozo V., Minguez A., Arrieta I., Cortegano I.,

RA Cardaba B., Gallardo S., Rojo M., Palomino P., Lahoz C.;

RT Molecular cloning of major allergen from Cupressus arizonica pollen:

RT Cup a 1.;

RT Clin. Exp. Allergy 30:1750-1758(2000).

RL -1 SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

CC AMB A I/AMB A II/CRY J I SUBFAMILY.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match: 100.0%; Score 78; DB 1; Length 346;  
Best Local Similarity: 100.0%; Pred. No. 4.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 211 KSMKVTVAENQFGPN 225

QY 1 KSMKVTVAENQFGPN 15

MPAL\_JUNAS STANDARD; PRT; 367 AA.

ID MPAL\_JUNAS STANDARD; PRT; 367 AA.

AC P81254; OSZNU7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DR Major pollen allergen Jun a 1 precursor.  
OS Uniiperus ashei (Ozark white cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferales; Coniferales; Cupressaceae; Uniiperus.  
OX NCBI\_TaxId=13101;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-38; 42-50; 58-80; 88-94;  
RP 117-124; 134-140; 160-164; 256-263 AND 322-325.  
RC MEDLINE=99414163; PubMed=10482836;  
RA Midoro-Horinuti T.M., Goldblum R.W., Kurosky A., Wood T.G.,  
RT Brooks E.G.;  
RT "Molecular cloning of mountain cedar ('Uniiperus ashei) pollen major  
RT allergen, Jun a 1";  
RL J. Allergy Clin. Immunol. 104:613-617(1999).  
[2]  
RP SEQUENCE OF 22-50.  
RP TISSUE=Pollen;  
RX MEDLINE=99414162; PubMed=10482835;  
RA Midoro-Horinuti T., Goldblum R.W., Kurosky A., Goetz D.W.,  
RA Brooks E.G.;  
RT "Isolation and characterization of the mountain cedar ('Uniiperus  
RT ashei) pollen major allergen, Jun a 1";  
RL J. Allergy Clin. Immunol. 104:608-612(1999).  
CC -1- DISEASE: THIS PROTEIN CAUSES SEVERAL SEASONAL ALLERGIC RHINITIS  
CC IN NORTH AMERICA.  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
AM A I/MB A II/CRY J I SUBFAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
-----

CC CC  
CC EMBL; AF106663; AAD03609.1; -.  
DR EMBL; AF106662; AAD03608.1; -.  
DR InterPro; IPRO02022; Amb allergen.  
DR Pfam; PF00544; pec lyase1.1.  
KW PRINTS; PR00807; AMBALLERGEN.  
DR Allergen; Glycoprotein; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 367 MAJOR POLLEN ALLERGEN JUN A 1.  
FT CARBOHYD 148 148 N-LINKED (GLCNAC . . . ) (POTENTIAL).  
FT CARBOHYD 178 178 N-LINKED (GLCNAC . . . ) (POTENTIAL).  
SQ SEQUENCE 367 AA; 39824 MW; FCSB81BE75662E49 CRC64;

Query Match 100.0%; Score 78; DB 1; Length 367;  
Best Local Similarity 100.0%; Pred. No. 4.6e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 1 KSMKVTVAFNQFGPN 15  
ID |||||||  
AC SBP CRYJA STANDARD; PRT; 374 AA.  
AD P18632;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Sugil basic protein precursor (SBP) (Major allergen Cry j 1) (Cry j I).  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferales; Coniferales; Cupressaceae; Cryptomeria.  
OX NCBI\_TaxId=3369;  
RN [1]  
RP SEQUENCE \*FROM N.A., AND PARTIAL SEQUENCE.



FT VARIANT 202 202 S -> T (IN CRY J 1-B).  
 FT VARIANT 221 221 L -> S (IN CRY J 1-B).  
 FT VARIANT 358 358 Q -> H (IN CRY J 1-B).  
 FT VARIANT 361 361 K -> O (IN CRY J 1-B).  
 SQ SEQUENCE 374 AA, 40645 MW, 74AB2550248F66F CRC64;

Query Match 100.0%; Score 78; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTAFAFGPN 15  
 DB 232 KSMKVTAFAFGPN 246

RESULT 4  
 MPAL\_CHA0B STANDARD; PRT; 375 AA.

AC Q96385; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major pollen allergen Cha o 1 precursor.  
 OS Chamaecyparis obtusa (Japanese cypress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Coniferales; Cupressaceae;  
 OC Chamaecyparidaceae.  
 NC NCBI\_TaxID=13415;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=pollen;  
 RX MEDLINE=96265194; Pubmed=8676896;  
 RA Suzuki M., Komiyama N., Itoh H., Some T., Kuno K., Takagi I.,  
 RA Ohta N.;  
 RT "Purification, characterization and molecular cloning of Cha o 1, a  
 RT major allergen of Chamaecyparis obtusa (Japanese cypress) pollen.";  
 RL Mol. Immunol. 33:451-460(1996).  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A I/AMB A II/CRY J I SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL; D45404; BAA08246.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR Allergoh; Glycoprotein; Signal.  
 FT CHAIN 1 21 MAJOR POLLEN ALLERGEN CHA O 1.  
 FT SIGNAL 1 21  
 FT CARBOHYD 22 375 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 375 AA; 40258 MW; 81CD91DF7066BBF CRC64;

Query Match 100.0%; Score 78; DB 1; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTAFAFGPN 15  
 DB 232 KSMKVTAFAFGPN 246

RESULT 5  
 9612\_LYCES

ID 9612 LYCES STANDARD; PRT; 404 AA.  
 AC P24395;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Style development-specific protein 9612 precursor.  
 GN 9612.

OS Lycopersicon esculentum (tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 NC NCBI\_TaxID=4081;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. VF36; TISSUE=Pistil;  
 RX MEDLINE=9117185; Pubmed=227637;  
 RA Badellier K.A., Smith A.G., Gasser C.S.;  
 RT "Regulation of a stylar transmitting tissue-specific gene in  
 RT wild-type and transgenic tomato and tobacco."  
 RL Mol. Gen. Genet. 224:183-192(1990).

CC -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPMENT OF THE TRANSMITTING  
 CC TISSUE OF THE STYLE AND/OR IN THE EVENTS RELATED TO POLLINATION  
 CC SUCH AS SOME ASPECT IN THE FACILITATION OF COMPATIBLE POLLEN TUBE  
 CC GROWTH.

CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN THE PISTIL WHERE IT IS  
 CC FOUND IN THE OUTER FIVE LAYERS OF THE STRANDS OF TRANSMITTING  
 CC TISSUE WITHIN THE UPPER TWO-THIRDS OF THE STYLE. FOUND AT MUCH  
 CC LOWER LEVELS IN THE ANTERS AND VEGETATIVE ORGANS.  
 CC -1- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND DURING  
 CC -1- SIMILARITY: 54% IDENTICAL TO TOMATO PROTEINS P59 (AC P15722)  
 CC AND P56 (AC P15721).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL; X55193; CAA38979.1; -  
 DR PIR; S12209; S12209.  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Signal.  
 FT CHAIN 1 20 POTENTIAL.  
 FT SIGNAL 1 20  
 FT CARBOHYD 21 404 STYLE DEVELOPMENT-SPECIFIC PROTEIN 9612.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 404 AA; 44298 MW; B26ED69B128D8675 CRC64;

Query Match 66.7%; Score 52; DB 1; Length 404;  
 Best Local Similarity 76.9%; Pred. No. 0.036;  
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KSMKVTAFAFGPN 13  
 DB 262 KSMKVTAFAFGPN 274

RESULT 6  
 PEL\_TOBAC STANDARD; PRT; 397 AA.  
 AC P40572;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Pectate lyase precursor (EC 4.2.2.2).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OC NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A. TISSUE=Pollen;  
 RC STRAIN=cv. Samsun; PubMed=1421152;  
 RA MEDLINE=93043039; Harvey A., Lonsdale D.M.;  
 RT "Isolation and characterization of a tobacco gene with homology to  
 RT pectate lyase which is specifically expressed during  
 RT microsporogenesis";  
 RL Plant Mol. Biol. 20:493-502(1992).  
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC -1- DEVELOPMENTAL STAGE: MAXIMALLY EXPRESSED LATE IN POLLEN  
 CC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X67158; CAA47630.1; -  
 DR EMBL; X67159; CAA47631.1; -  
 DR EMBL; X61102; CAA43414.1; -  
 DR PIR; S26211; S26211.  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR Lyase; Signal.  
 FT CHAIN 1 25  
 FT SIGNAL 26 397  
 FT ACT\_SITE 272 272  
 FT CARBOHYD 134 134  
 FT CARBOHYD 227 227  
 FT CONFLICT 156 156  
 FT CONFLICT 189 190  
 FT CONFLICT 200 200  
 FT CONFLICT 202 202  
 FT CONFLICT 249 249  
 FT SEQUENCE 397 AA; 44351 MW; EF0A82C5DA7643F CRC64;  
 SQ  
 Query Match 62.8%; Score 49; DB 1; Length 397;  
 Best Local Similarity 61.5%; Pred. No. 0.13;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 KSMKTVAFNPG 13  
 Db 254 KDMKITLVNHPG 266  
 ID PEL\_LILLO STANDARD; PRT; 434 AA.  
 AC P40973;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Pectate lyase precursor (EC 4.2.2.2).  
 OS Lilium longiflorum (Trumpet Lily).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae;  
 OC Lilium.  
 OC NCBI\_TaxID=4690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nellie white; TISSUE=Pollen;  
 RA Kim S.R., Finkel D.J., An G.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z17328; CAA78976.1; -  
 DR EMBL; L18911; AAA33398.1; -  
 DR PIR; S29612; S29612.  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 DR Lyase; Signal.  
 FT CHAIN 1 22  
 FT SIGNAL 23 434  
 FT ACT\_SITE 312 312  
 FT CARBOHYD 68 68  
 FT CARBOHYD 97 97  
 FT SEQUENCE 434 AA; 48457 MW; C1F3E30AD2BBD064 CRC64;  
 SQ  
 Query Match 60.3%; Score 47; DB 1; Length 434;  
 Best Local Similarity 81.8%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 MKVTVAFNPG 13  
 Db 296 MQVTVAFNHPG 306  
 ID PEL\_LILLO STANDARD; PRT; 398 AA.  
 AC P15721;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Probable pectate lyase P56 precursor (EC 4.2.2.2).  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 OC NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. VF36; TISSUE=Anther;  
 RA MEDLINE=91322485; PubMed=1983191;  
 RA Wing R.A., Yamaguchi J., Latabell S.K., Urein V.M., McCormick S.;  
 RT "Molecular and genetic characterization of two pollen-expressed genes  
 RT that have sequence similarity to pectate lyases of the plant pathogen  
 RT Erwinia";  
 RL Plant Mol. Biol. 14:117-28(1990).  
 CC [2]  
 CC REVISIONS.  
 CC STRAIN=cv. VF36; TISSUE=Anther;  
 CC Wing R.A.;  
 CC Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND TUBE  
 CC GROWTH.  
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X15500; CAA33524.1; -  
DR PIR; S08383; S08383.  
DR InterPro; IPR002022; Amb allergen.  
DR Pfam; PF00544; pec lyase; 1.  
DR PRINTS; PR00807; AMBALLERGEN.  
KW Lyase; Multigene family; Signal.  
FT SIGNAL 1 27 OR 22 (POTENTIAL).  
FT CHAIN 28 398 PROBABLE PEPTIDE LYASE P56.  
FT ACT SITE 273 273 POTENTIAL.  
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 398 AA; 44563 MW; 8D676250BD8C7C8 CRC64;

Query Match 59.0%; Score 46; DB 1; Length 398;  
Best Local Similarity 53.8%; Pred. No. 0.47;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKTVAFNORG 13  
Db 255 KSMKTVAFNORG 267

## RESULT 9

PE59 LYCES STANDARD; PRT; 449 AA.  
AC P15722;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Probable peptidase lyase P59 precursor (EC 4.2.2.2).  
GN LAPT59.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID:4081;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=cv. VF36; TISSUE=anther;  
RA MEDLINE=91322485; PubMed=1983191;  
RT Wang R.A., Yamaguchi J., Larabell S.K., Ursin V.M., McCormick S.;  
RT "Molecular and genetic characterization of two pollen-expressed genes  
RT that have sequence similarity to peptidase lyases of the plant pathogen  
RT Brevintra";  
RT Plant Mol. Biol. 14:17-28(1990).

CC -1- FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND TUBE  
CC GROWTH.  
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of peptate to give  
CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
CC their non-reducing ends.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X15499; CAA33523.1; -  
DR PIR; S27098; S27098.  
DR InterPro; IPR002022; Amb allergen.  
DR Pfam; PF00544; pec lyase; 1.  
DR PRINTS; PR00807; AMBALLERGEN.

KW Lyase; Multigene family; Signal.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 449 PROBABLE PEPTIDE LYASE P59.  
FT ACT SITE 325 325 POTENTIAL.  
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 449 AA; 50893 MW; 17E3AA13F173B03C CRC64;

Query Match 55.1%; Score 43; DB 1; Length 449;  
Best Local Similarity 63.6%; Pred. No. 1.9;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 MKVTVAFNORG 13  
Db 309 MKVTVAFNORG 319

## RESULT 10

MP13 AMBAR STANDARD; PRT; 397 AA.  
AC P27761;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pollen allergen Amb A 1.3 precursor (Antigen E) (Antigen Amb A 1).  
OS Ambrosia artemisiifolia (Short ragweed).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
OC Heliantheae; Ambrosia.  
OX NCBI\_TaxID:4212;

RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Pollen;  
RA MEDLINE=91093235; PubMed=1702434;  
RT Rainar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,  
RT Klapper D.G.;  
RT "Cloning of Amb A I (antigen E), the major allergen family of short  
RT ragweed pollen";  
RT J. Biol. Chem. 266:1229-1236(1991).

RN [2]  
RP SEQUENCE FROM N.A., AND VARIANTS.  
RC TISSUE=Pollen;  
RA MEDLINE=92234570; PubMed=1809687;  
RT Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.;  
RT "Sequence polymorphism of Amb A I and Amb A II, the major allergens  
RT in Ambrosia artemisiifolia (short ragweed).";  
RT Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).

CC -1- SUBUNIT: MONOMER.  
CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
CC -1- PTM: The N-terminus is blocked.  
CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

CC AMB A I/AMB A II/CRY J I SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M62961; AAA32668.1; -  
DR EMBL; M60360; AAA32669.1; ALT\_SEQ.  
DR PIR; C39099; C39099.  
DR InterPro; IPR002022; Amb allergen.  
DR Pfam; PF00544; pec lyase; 1.  
DR PRINTS; PR00807; AMBALLERGEN.  
KW Antigen; Allergen; Signal; Multigene family; Polymorphism.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 397 POLLEN ALLERGEN AMB A 1.3.

FT VARIANT 48 48 L -> Y.  
SQ SEQUENCE 397 AA; 42928 MM; C8DB4125/590DD0A CRC64;

Query Match 53.8%; Score 42; DB 1; Length 397;  
Best Local Similarity 60.0%; Pred. No. 2.6;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
DB 256 KGMATVAFNMTDN 270

RESULT 11

IFU SYNY3 STANDARD; PRT; 1199 AA.  
AC P52965;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative pyruvate-flavodoxin oxidoreductase (EC 1.2.7.-).  
GN NIFU OR SL0741.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N., Sugita M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 644 to 924 of the genome.";  
RL DNA Res. 2:153-166(1995).  
CC -1- FUNCTION: OXIDOREDUCTASE REQUIRED FOR THE TRANSFER OF ELECTRONS FROM PYRUVATE TO FLAVODOXIN (POTENTIAL).  
CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + oxidized flavodoxin = acetyl-CoA + CO(2) + reduced flavodoxin.  
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D64005; BAA10774.1; -;  
DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
DR InterPro; IPR002869; POR.  
DR InterPro; IPR002860; POR\_N.  
DR Pfam; PF00037; Fer4; 2.  
DR Pfam; PF01558; POR; 1.  
DR Pfam; PF01855; POR\_N; 1.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
KW Oxidoreductase; Nitrogen fixation; Electron transport; Iron-sulfur; 4Fe-4S; Complete proteome.  
KW METAL 690  
FT METAL 693  
FT METAL 693  
FT METAL 696  
FT METAL 700  
FT METAL 746  
FT METAL 746  
FT METAL 749  
FT METAL 752  
FT METAL 756  
SQ SEQUENCE 1199 AA; 131457 MM; 12P5C34CB8BD50B6 CRC64;

Query Match 53.8%; Score 42; DB 1; Length 1199;  
Best Local Similarity 60.0%; Pred. No. 7.9;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 KSMKVTVAFNQFGPN 15

DB 459 KSGSVTVSHRFGPN 473

RESULT 12

G3PA GRAVE STANDARD; PRT; 416 AA.  
AC G3PA GRAVE  
ID G3PA GRAVE  
AD P30724;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glyceralddehyde 3-phosphate dehydrogenase, chloroplast precursor (EC 1.2.1.13) (NADP-dependent glyceralddehyde phosphate dehydrogenase).  
GN GABA.  
OS Gracilaria verrucosa.  
OC Eukaryota; Rhodophyta; Florideophyceae; Gracilariaceae; Gracilariaceae.  
OX NCBI\_TaxID=2777;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93306755; PubMed=7916671;  
RA Zhou Y.H., Ragan M.A.;  
RT "Cloning and characterization of the nuclear gene encoding plastid glyceralddehyde-3-phosphate dehydrogenase from the marine red alga Gracilaria verrucosa.";  
RT Curr. Genet. 23:483-489(1993).  
RL [2]  
RN [2]  
RP REVISIONS.  
RX MEDLINE=95042840; PubMed=7954900;  
RA Zhou Y.H., Ragan M.A.;  
RT "Cloning and characterization of the nuclear gene encoding plastid glyceralddehyde-3-phosphate dehydrogenase from the marine red alga Gracilaria verrucosa.";  
RT Curr. Genet. 26:79-86(1994).  
RL Curr. Genet. 26:79-86(1994).  
CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate + NADP(+) = 3-phospho-D-glyceroyl phosphate + NADPH.  
CC -1- PATHWAY: Calvin cycle.  
CC -1- SUBUNIT: Homotrimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z15102; CAA78811.1; -;  
DR EMBL; L22011; AAA33355.1; -;  
DR PIR; S25596; S25596.  
DR PIR; S45484; S45484.  
DR HSP; P50362; INMG.  
DR InterPro; IPR000173; GAP\_dhrogenase.  
DR Pfam; PF00044; gpdh; 1.  
DR Pfam; PF02800; gpdh\_C; 1.  
DR PRINTS; PR00078; G3PDHGRNAS.  
DR PROSITE; PS00071; GAPDH; 1.  
KW Calvin cycle; Oxidoreductase; NADP; Transit peptide; Chloroplast; Multigene family.  
KW TRANSIT 1  
FT CHAIN 79  
FT BINDING 233  
FT ACT\_SITE 260  
SQ SEQUENCE 416 AA; 44337 MM; 33B9437B84230BD8 CRC64;

Query Match 52.6%; Score 41; DB 1; Length 416;  
Best Local Similarity 75.0%; Pred. No. 4.2;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 SKMYTVAFNQFG 13  
DB 78 SKMYRVAINQFG 89

RESULT 13  
EPH2\_HUMAN STANDARD; PRT; 333 AA.  
AC P52759;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)  
DE (LEK-5) (HTK ligand) (HTK-L).  
GN EPHB2 OR EPLG5 OR LERKS OR HTKL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC- TISSUE=Brain;  
RX MEDLINE=96145238; PubMed=8559144;  
RA Cerretti D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,  
RA Maraskovsky E., Park L.S., Lyman S.D., Copeland N.G., Gilbert D.J.,  
RA Jenkins N.A., Fletcher R.A.,  
RT "Isolation of LERK-5: a ligand of the eph-related receptor tyrosine  
RT kinases.";  
RL Mol. Immunol. 32:1197-1205(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC- TISSUE=Brain;  
RX MEDLINE=95199254; PubMed=7534404;  
RA Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D.,  
RA Gillett N., Matthews W.,  
RT "Molecular cloning of a ligand for the EPH-related receptor protein-  
RT tyrosine kinase Htk.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98192220; PubMed=9533549;  
RA Vogt T., Stoltz W., Welsh J., Jung B., Kerbel R.S., Kobayashi H.,  
RA Landthaler M., McClelland M.,  
RT "Overexpression of lerk-5/epig5 messenger RNA: a novel marker for  
RT increased tumorigenicity and metastatic potential in human malignant  
RT melanomas.";  
RL Clin. Cancer Res. 4:791-797(1998).  
RN [4]  
RP FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF  
CC LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY). EPHB4 AND EPHA3.  
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB4 AND EPHA3.  
CC -1- TISSUE SPECIFICITY: LONG AND KIDNEY.  
CC -1- PFM: INDICIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE  
CC CYTOPLASMIC DOMAIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DB EMBL; U16797; AAA9707.1; -;  
DB EMBL; U16797; AAC41752.1; -;  
DB EMBL; U81262; AAD03786.1; -;  
DB Genew; HGNC:3227; EPHB2.  
DB MIM; 600527; -;  
DB InterPro; IPR001799; Ephrin.  
DB Pfam; PF00812; Ephrin; 1.  
DB PRINTS; PR01347; EPHRIN.  
DB ProDom; PD002533; Ephrin; 1.

DR PROSITE; PS01299; EPHRIN; 1.  
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
FT SIGNAL; Phosphorylation.  
FT CHAIN 1 27 POTENTIAL.  
FT CHAIN 28 333 EPHRIN-B2.  
FT DOMAIN 28 229 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 230 250 POTENTIAL.  
FT DOMAIN 251 333 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 331 333 PDZ RECOGNITION MOTIF (POTENTIAL).  
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 333 AA; 36923 MW; 6D9932A632626A9A CRC64;

Query Match 51.3%; Score 40; DB 1; Length 333;  
Best local similarity 40.0%; Pred. No. 5.2;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 SKMYTVAFNQFGPN 15  
DB 106 QDIKTIKQFERSPN 120

RESULT 14  
EFB2\_MOUSE STANDARD; PRT; 336 AA.  
ID EFB2\_MOUSE  
AC P52800;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)  
DE (LEK-5) (HTK ligand) (HTK-L) (ELF-2)  
GN EPHB2 OR EPLG5 OR LERKS OR HTKL OR ELF2 OR EPL5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96145238; PubMed=8559144;  
RA Cerretti D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,  
RA Maraskovsky E., Park L.S., Lyman S.D., Copeland N.G.,  
RA Gilbert D.J., Jenkins N.A., Fletcher R.A.,  
RT "Isolation of LERK-5: a ligand of the eph-related receptor tyrosine  
RT kinases.";  
RL Mol. Immunol. 32:1197-1205(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB57BL/6J X SJL/J;  
RX MEDLINE=95199254; PubMed=7534404;  
RA Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D.,  
RA Gillett N., Matthews W.,  
RT "Molecular cloning of a ligand for the EPH-related receptor protein-  
RT tyrosine kinase Htk.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ICR; TISSUE=Brain;  
RX MEDLINE=95379837; PubMed=7651410;  
RA Bergmann A.D., Cheng H.J., Brambilla R., Klein R., Planagan J.G.,  
RT "ELF-2, a new member of the Eph ligand family, is segmentally  
RT expressed in mouse embryos in the region of the hindbrain and newly  
RT forming somites.";  
RL Mol. Cell. Biol. 15:4921-4929(1995).  
RN [4]  
RP FUNCTION.  
RN MEDLINE=20171264; PubMed=10704386;  
RA Imondi R., Wideman C., Kaprielian Z.,  
RT "Complementary expression of transmembrane ephrins and their receptors  
RT in the mouse spinal cord: a possible role in constraining the  
RT orientation of longitudinally projecting axons.";  
RL Development 127:1397-1410(2000).  
CC -1- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF  
CC LONGITUDINALLY PROJECTING AXONS.

CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS, SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE PERIOD OF COMMISSURAL AXON PATHFINDING.  
 CC -1- PFM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE CYTOPLASMIC DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U16819; AAA9708.1; -  
 DR EMBL: U38847; AAC42052.1; -  
 DR EMBL: U30244; AAA82934.1; -  
 DR MGI: 105097; Efb2.  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.  
 DR PRINTS: PR01347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin.  
 DR PROSITE: PS01299; EPHRIN.  
 DR Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
 KW Signal; Phosphorylation.  
 FT CHAIN 1 28  
 FT SIGNAL 1 28  
 FT DOMAIN 29 232  
 FT TRANSMEM 233 253  
 FT DOMAIN 254 336  
 FT DOMAIN 337 399  
 FT CARBOHYD 39 39  
 FT CARBOHYD 142 142  
 FT CONFLICT 3 4  
 FT CONFLICT 177 177  
 FT CONFLICT 177 177  
 SQ SEQUENCE 336 AA; 37202 MW; D0889499639554 CRC64;  
 Query Match 51.3%; Score 40; DB 1; Length 336;  
 Best Local Similarity 40.0%; Pred. No. 5.2;  
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQPGN 15  
 Db 109 QDVKFTIKFOEFSN 123  
 RESULT 15  
 EFB1\_XENLA  
 ID EFB1\_XENLA STANDARD; PRT; 327 AA.  
 AC 013057;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DT Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)  
 DE (IERK-2) (ELK ligand) (ELK-L) (XLERK).  
 GN EPHB1 OR EPIC2 OR IERK2.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97316777; PubMed=9174051;  
 RA Jones T.L., Katavanova I., Chong L., Zhou R.P., Daar I.O.;  
 RT "Identification of Xlerk, an Eph family ligand regulated during mesoderm induction and neurogenesis in Xenopus laevis.";

RL Oncogene 14:2159-2166 (1997).  
 CC -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPING MESENCHYMAL AND NERVOUS TISSUE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT LOW LEVELS IN MOST ADULT TISSUES WITH INCREASED LEVELS OBSERVED IN THE KIDNEY, OOCYTES, OVARY AND TESTIS.  
 CC -1- PFM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE CYTOPLASMIC DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U31427; AAC35995.1; -  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.  
 DR PRINTS: PR01347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin.  
 DR PROSITE: PS01299; EPHRIN.  
 DR Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
 KW Signal; Phosphorylation.  
 FT CHAIN 1 20  
 FT SIGNAL 1 20  
 FT CHAIN 21 327  
 FT DOMAIN 21 225  
 FT TRANSMEM 226 246  
 FT DOMAIN 247 327  
 FT DOMAIN 328 327  
 FT CARBOHYD 131 131  
 FT CARBOHYD 202 202  
 SQ SEQUENCE 327 AA; 36621 MW; 71230CE7F6B5974 CRC64;  
 Query Match 50.0%; Score 39; DB 1; Length 327;  
 Best Local Similarity 40.0%; Pred. No. 7.8;  
 Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQPGN 15  
 Db 98 KEVRFITIKFOEFSN 112  
 Search completed: April 20, 2003, 13:07:37  
 Job time: 3.92105 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 / Search time 12.6711 Seconds

(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524d-57

Perfect score: 78

Sequence: 1 KSMKVTVAFNQFGPN 15

Scoring table: BLOSUM62

Searched: Gapop 10.0, Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

- 1: SPTREMBL 21:\*
- 2: sp\_archaea:\*
- 3: sp\_bacteria:\*
- 4: sp\_fungi:\*
- 5: sp\_human:\*
- 6: sp\_invertebrate:\*
- 7: sp\_mammal:\*
- 8: sp\_mhc:\*
- 9: sp\_organelle:\*
- 10: sp\_phage:\*
- 11: sp\_plant:\*
- 12: sp\_rodent:\*
- 13: sp\_virus:\*
- 14: sp\_vertebrate:\*
- 15: sp\_unclassified:\*
- 16: sp\_virus:\*
- 17: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	367	10	Q91LT2
2	78	100.0	367	10	Q91LT1
3	78	100.0	367	10	Q9M4S6
4	78	100.0	367	10	Q9M4S5
5	78	100.0	367	10	Q9M4S4
6	78	100.0	367	10	Q9M4S3
7	78	100.0	367	10	Q9M4S2
8	78	100.0	367	10	Q9M4S1
9	78	100.0	367	10	Q9M4S0
10	78	100.0	367	10	Q9M4S0
11	78	100.0	367	10	Q9M4S0
12	78	100.0	367	10	Q9M4S0
13	78	100.0	367	10	Q9M4S0
14	78	100.0	367	10	Q9M4S0
15	78	100.0	367	10	Q9M4S0
16	78	100.0	367	10	Q9M4S0

17	55	70.5	407	10	Q9SDW4	Q9sdw4 musa acumin
18	52	66.7	226	10	Q23667	Q23667 arabidopsis
19	52	66.7	392	10	Q9FXD8	Q9fxd8 arabidopsis
20	52	66.7	398	10	Q9M505	Q9m505 vitis vinif
21	52	66.7	401	10	Q24554	Q24554 zinnia eleg
22	52	66.7	404	10	Q9S871	Q9s871 arabidopsis
23	52	66.7	408	10	Q8W116	Q8w116 arabidopsis
24	52	66.7	409	10	Q91T20	Q91t20 arabidopsis
25	52	66.7	410	10	Q93XJ1	Q93xj1 salix gligi
26	52	66.7	410	10	Q9FMK5	Q9fmk5 arabidopsis
27	52	66.7	432	10	Q93Z25	Q93z25 arabidopsis
28	51	65.4	324	10	Q8S345	Q8s345 capsicum an
29	51	65.4	349	10	Q93Z77	Q93z77 arabidopsis
30	51	65.4	455	10	Q64510	Q64510 arabidopsis
31	51	65.4	459	10	Q23665	Q23665 arabidopsis
32	51	65.4	459	10	Q9M9S2	Q9m9s2 arabidopsis
33	50	64.1	408	10	Q9CSM8	Q9csm8 arabidopsis
34	49	62.8	263	10	Q04084	Q04084 arabidopsis
35	49	62.8	354	10	Q94001	Q94001 arabidopsis
36	49	62.8	431	10	Q23017	Q23017 arabidopsis
37	49	62.8	450	10	Q40319	Q40319 arabidopsis
38	49	62.8	475	10	Q9S8H4	Q9s8h4 arabidopsis
39	48	61.5	143	10	Q94B14	Q94b14 vitis vinif
40	48	61.5	368	10	Q94FT5	Q94ft5 fragaria an
41	48	61.5	392	10	Q9FM66	Q9fm66 arabidopsis
42	48	61.5	405	10	Q24416	Q24416 fragaria an
43	48	61.5	405	10	Q94FT6	Q94ft6 fragaria an
44	48	61.5	452	10	Q91RMS	Q91rms arabidopsis
45	48	61.5	472	10	Q91RFS	Q91rfs arabidopsis

## ALIGNMENTS

## RESULT 1

ID Q91LT2 PRELIMINARY; PRT; 367 AA.

AC Q91LT2;

DT 01-OCT-2000 (TREMREL. 15, Created)

DT 01-OCT-2000 (TREMREL. 15, Last sequence update)

DT 01-JUN-2002 (TREMREL. 21, Last annotation update)

DE Pollen major allergen 1-2.

OS Juniperus virginiana (Eastern red cedar).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferophyta; Coniferales; Cupressaceae; Juniperus.

OX NCBI\_TaxID=39584;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21315424; PubMed=11422137;

RA Midoro-Horizuti T.M., Goldblum R.M., Brooke E.G.;

RT "Identification of mutations in the genes for the pollen allergens of

RT eastern red cedar (Juniperus virginiana).";

RL Clin. Exp. Allergy 31:771-778(2001).

DR EMBL, AF154427; AAF80164.1; -

DR InterPro; IPR02022; Amb allergen.

DR Pfam; PF00544; pec lyase; 1.

DR PRINTS; PR00807; AMBALLERGEN.

SQ SEQUENCE 367 AA; 39768 MW; 0A6AC2F1BAF89586 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;

Best Local Similarity 100.0%; Pred. No. 4e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15

DB 232 KSMKVTVAFNQFGPN 246

## RESULT 2

ID Q91LT1 PRELIMINARY; PRT; 367 AA.

AC Q91LT1;

DT 01-OCT-2000 (TREMREL. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Pollen major allergen 1.1.  
 OS Juniperus virginiana (Eastern red cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxID=39584;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21315424; PubMed=11422137;  
 RA Miodoro-Horiet T.M., Goldblum R.M., Brooks E.G.;  
 RT "Identification of mutations in the genes for the pollen allergens of  
 eastern red cedar (Juniperus virginiana).";  
 RL Clin. Allergy 31:771-778 (2001).  
 DR EMBL; AF151429; AAF80166.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39708 MW; DCBD1981A74E4711 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 3

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

AC Q9M4S6 PRELIMINARY; PRT; 367 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP51.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257491; AAF72625.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39835 MW; B4E9C60108C2C5A3 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 4

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

AC Q9M4S5 PRELIMINARY; PRT; 367 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP51.  
 OS Cupressus sempervirens.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257492; AAF72626.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39894 MW; 5D56FC0E3263B741 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 5

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

AC Q9M4S4 PRELIMINARY; PRT; 367 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP51.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257493; AAF72627.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39934 MW; 974D3011D74B3D6E CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 6

OY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

AC Q9M4S3 PRELIMINARY; PRT; 367 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP51.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxID=13469;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RA Monsalve R.I., Villalba M., Rodriguez R.;  
 RT "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; Pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39832 MW; B5DFB5A61C07A53 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 7

Q9M4S2 PRELIMINARY; PRT; 367 AA.  
 ID Q9M4S2  
 AC Q9M4S2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cup s 1 pollen allergen.  
 GN CUP81.  
 OS Cupressus sempervirens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=13469;

RA "Cloning and expression of Cup s 1, the major allergen of the pollen  
 of Cupressus sempervirens.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF257494; AAF72628.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; Pec\_lyase; 1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 367 AA; 39819 MW; AE7E055A61C07A53 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 8

Q93X51 PRELIMINARY; PRT; 367 AA.  
 ID Q93X51  
 AC Q93X51;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Putative allergen jun o 1.  
 GN JUN O 1.  
 OS Juniperus oxycedrus (Prickly juniper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.  
 OX NCBI\_TaxId=99008;

RA "Cloning of Juniperus oxycedrus major allergen.";  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ293767; CAC8400.1; -

DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; Pec\_lyase; 1.  
 SQ SEQUENCE 367 AA; 39808 MW; 5D28204DBFD1B9D7 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 9

Q8RUR1 PRELIMINARY; PRT; 374 AA.  
 ID Q8RUR1  
 AC Q8RUR1;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cry J 1 precursor.  
 GN Cry J 1.1 OR Cry J 1.2.  
 OS Cryptomeria japonica (Japanese cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
 OX NCBI\_TaxId=3369;

RA "Isolation and characterization of cDNAs encoding major allergen Cry J  
 1 from Cryptomeria japonica pollen.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB081309; BAB86286.1; -  
 DR EMBL; AB081310; BAB86287.1; -  
 KW SIGNAL.  
 FT CHAIN 1 21 POTENTIAL.  
 SQ SEQUENCE 374 AA; 40720 MW; 90D0085D24BF2BD4 CRC64;

Query Match 100.0%; Score 78; DB 10; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
 DB 232 KSMKVTVAFNQFGPN 246

## RESULT 10

Q93XL6 PRELIMINARY; PRT; 367 AA.  
 ID Q93XL6  
 AC Q93XL6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative allergen Cup a 1 precursor.  
 GN CUP A 1.  
 OS Cupressus arizonica.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cupressus.  
 OX NCBI\_TaxId=49011;

RA "Cloning of Cupressus arizonica major allergen.";  
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ278498; CAC37790.2; -  
 KW SIGNAL.  
 FT CHAIN 1 21 POTENTIAL.  
 SQ SEQUENCE 367 AA; 39809 MW; AAF97260423A9F28 CRC64;

Query Match 93.6%; Score 73; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SMKVTVAFNQFGPN 15  
 |||||  
 DB 233 SMKVTVAFNQFGPN 246

RESULT 11  
 Q9FY87 PRELIMINARY; PRT; 297 AA.  
 ID Q9FY87  
 AC Q9FY87  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Major pollen allergen-like protein.  
 GN T5E8 80.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL391712; CAC05454.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 SQ SEQUENCE 297 AA; 32971 MW; A75287A90A96D240 CRC64;

Query Match 79.5%; Score 62; DB 10; Length 297;  
 Best Local Similarity 73.3%; Pred. No. 0.0029;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
 |||||  
 DB 181 KDMKVTVAFNQFGPN 195

RESULT 12  
 Q9C8G4 PRELIMINARY; PRT; 368 AA.  
 ID Q9C8G4  
 AC Q9C8G4  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Pectate lyase, putative.  
 GN T4K22.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CV, COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Bugher E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressey T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gall J.E., Goldenfild A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.U., Koo H.L., Kremenskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Malt R., Marzalli A.,  
 RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizko M., Rooney T., Rowley D.,  
 RA Sakano H., Salzer J.S., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.O., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utecherback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:616-620(2000).  
 DR EMBL; AC025295; AAG51103.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 368 AA; 40576 MW; F35EDED82CED34BC CRC64;

Query Match 74.4%; Score 58; DB 10; Length 368;  
 Best Local Similarity 71.4%; Pred. No. 0.02;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGP 14  
 |||||  
 DB 229 KDMKVTVAFNQFGP 242

RESULT 13  
 ID 065457 PRELIMINARY; PRT; 394 AA.  
 AC 065457  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Pectate lyase like protein.  
 GN F1N20.190 OR ATG422090.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,  
 RA Schueller C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL022140; CAA18112.1; -  
 DR EMBL; AL161556; CAB79164.1; -  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase.1.  
 DR PRINTS; PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 394 AA; 43299 MW; 3912C1F6A8C6E3F5 CRC64;

Query Match 74.4%; Score 58; DB 10; Length 394;  
 Best Local Similarity 91.7%; Pred. No. 0.021;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 MKVTVAFNQFGP 14  
 |||||  
 DB 256 MKVTVAFNQFGP 267

RESULT 14

```

065456 ID 065456 PRELIMINARY; PRT; 394 AA.
AC 065456;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Pectate lyase like protein.
GN F12P1.180 OR AT4G22080.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Beyan M., Wedler H., Mambutt R., Bancroft I., Mewes H.W., Mayer K.,
RA Schueller C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Wedler H., Mambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022140; CA1811.1; -.
DR EMBL; AL161556; CAB79163.1; -.
DR InterPro; IPR002022; Amb_allergen.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00544; pec_lyase; 1.
DR PRINTS; PR00807; AMBALLERGEN.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW lyase.
SQ SEQUENCE 394 AA; 43476 MW; A48D586ECF148CA CRC64;

Query Match
Best Local Similarity 74.4%; Score 58; DB 10; Length 394;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 MKVTVAFNQFGP 14
DB 256 MKVTVAFNHFGP 267

RESULT 15
065388 ID 065388 PRELIMINARY; PRT; 390 AA.
AC 065388;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE F12P1.22 protein.
GN F12P1.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. COLUMBIA;
RA Vysotskaia V.S., Osborne B.I., Schwartz J.R., Toriumi M., Yu G.,
RA Kwan A., Oji O., Liu S., Buehler E., Conway A.B., Conway A.R.,
RA Dewar K., Feng J., Kim C., Kurtz D., Palm C.J., Li X., Shinn P.,
RA Sun H., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F12P1 sequence, complete
sequence."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002131; AAC17625.1; -.
DR InterPro; IPR002022; Amb_allergen.
DR Pfam; PF00544; pec_lyase; 1.
DR PRINTS; PR00807; AMBALLERGEN.

```

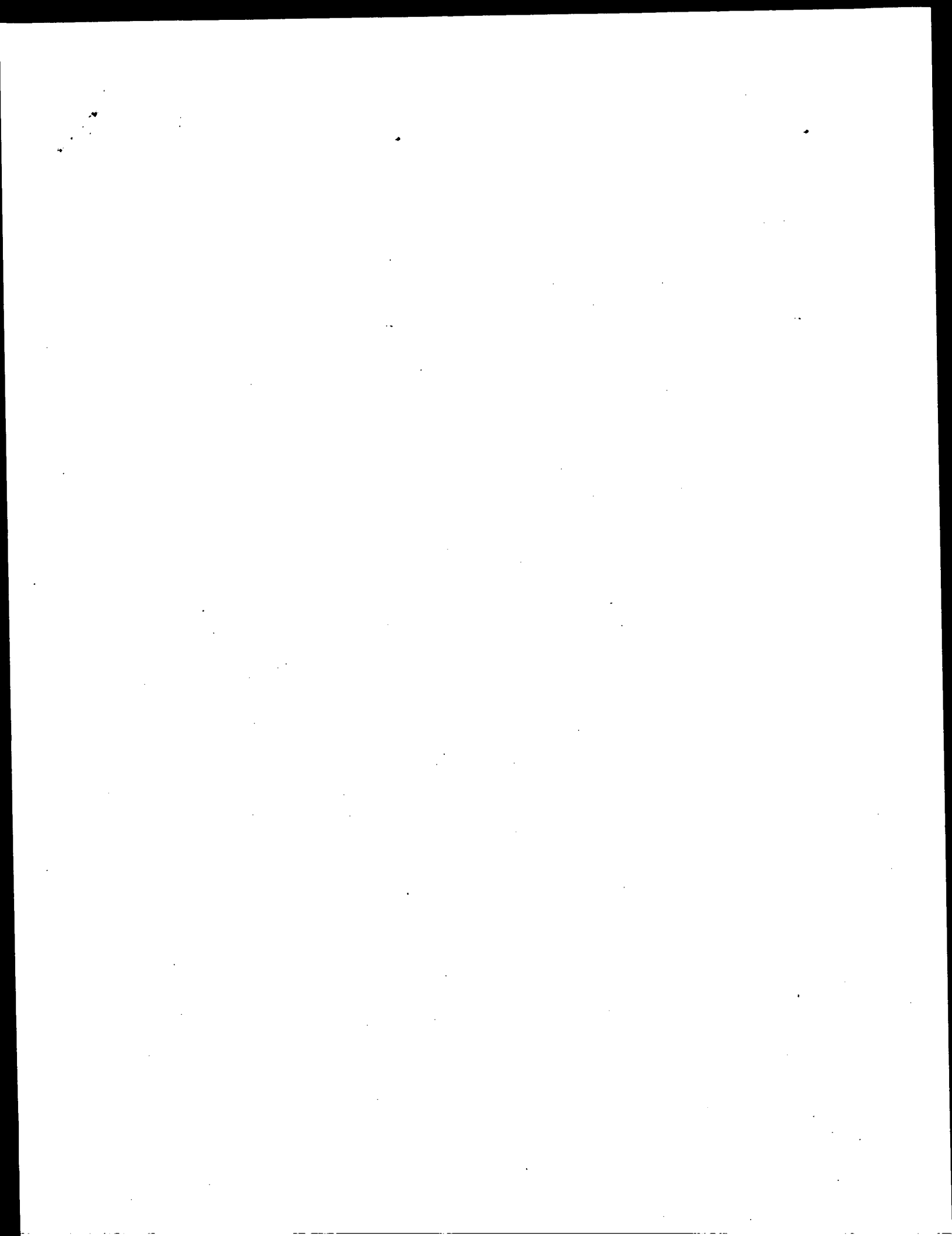
```

SQ SEQUENCE 390 AA; 43354 MW; 7E4D3CA9BB398CB7 CRC64;
Query Match
Best Local Similarity 73.1%; Score 57; DB 10; Length 390;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGP 14
DB 244 KMKVTVAFNHFGP 257

Search completed: April 20, 2003, 13:13:07
Job time : 13.6711 secs

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: us-09-142-524d-97

Perfect score: 81

Sequence: 1 VDGIIAAYQNPASWK 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5 ;

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 73:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	514	2	Cry j II protein -
2	81	100.0	514	2	second major aller
3	61	75.3	507	2	Jun a 2 protein -
4	48	59.3	514	2	polygalacturonase
5	45	55.6	374	2	hypothetical prote
6	45	55.6	1113	2	hypothetical prote
7	45	55.6	1175	2	hypothetical prote
8	45	55.6	1198	2	hypothetical prote
9	45	55.6	1201	2	hypothetical prote
10	45	55.6	1285	2	hypothetical prote
11	45	55.6	1305	2	hypothetical prote
12	45	55.6	1314	2	hypothetical prote
13	45	55.6	1756	2	hypothetical prote
14	44	54.3	521	2	probable matrix me
15	43	53.1	646	2	serine protein kin
16	43	53.1	1036	2	two-component sens
17	43	53.1	1102	2	methy1-accepting c
18	43	53.1	1102	2	two-component sens
19	43	53.1	1645	2	AG1897
20	42	51.9	1207	2	AG1897
21	42	51.9	1207	2	AG1897
22	42	51.9	1207	2	AG1897
23	42	51.9	1207	2	AG1897
24	42	51.9	1207	2	AG1897
25	41	50.6	327	2	hypothetical prote
26	41	50.6	327	2	hypothetical prote
27	40.5	50.0	544	2	phosphoglucosylase
28	40	49.4	94	2	hypothetical prote
29	40	49.4	94	2	hypothetical prote

30	40	49.4	200	2	D72389	conserved hypothe
31	40	49.4	212	2	J01378	glutathione transf
32	40	49.4	230	2	E70927	probable trnd prot
33	40	49.4	249	2	S75749	hypothetical prote
34	40	49.4	302	2	T40226	metotic recombinat
35	40	49.4	351	1	SAVZVR	hypothetical prote
36	40	49.4	351	1	SAVZVR	surface antigen pr
37	40	49.4	351	1	SAVZVR	surface antigen pr
38	40	49.4	351	1	SAVZVR	surface antigen pr
39	40	49.4	422	2	B75501	glycogen synthase
40	40	49.4	644	2	F82145	conserved hypothe
41	40	49.4	2348	2	AD1841	hypothetical prote
42	39.5	48.8	573	2	T23589	hypothetical prote
43	39	48.1	126	1	FAD01	profilin I - slime
44	39	48.1	127	2	B81799	diacylglycerol kin
45	39	48.1	127	2	C81069	diacylglycerol kin

#### ALIGNMENTS

RESULT 1  
S48730  
Cry j II protein - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999  
C/Accession: S48730  
R/Namba, M.; Kurose, M.; Toriige, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.; FERS Lett. 353, 124-128, 1994  
A/Title: Molecular cloning of the second major allergen, Cry j II, from Japanese cedar  
A/Reference number: S48730; MUID:95010777; PMID:7926035  
A/Accession: S48730  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-514 <NAM>  
A/Cross-References: GB:D37765; NID:9577695; PIDN:BAA07021.1; PID:dl007598; PID:9577696

Query Match 100.0%; Score 81; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAYQNPASWK 15  
Db 120 VDGIIAAYQNPASWK 134

RESULT 2  
JC2498  
second major allergen Cry j II precursor - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000  
C/Accession: JC2498; PC2346; A60147  
R/Komiyama, N.; Some, T.; Shimizu, K.; Morikubo, K.; Kino, K. Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994  
A/Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese cedar  
A/Reference number: JC2498; MUID:94271186; PMID:8002972  
A/Accession: JC2498  
A/Molecule type: mRNA  
A/Residues: 1-514 <KOM>  
A/Cross-References: DDBJ:D29772; NID:9506857; PIDN:BAA06172.1; PID:9506858  
A/Accession: PC2346  
A/Molecule type: protein  
A/Residues: 52-61 <KO2>  
R/Sakaguchi, M.; Inoue, S.; Tanai, M.; Ando, S.; Ueui, M.; Matuhasi, T. Allergy 45, 309-312, 1990  
A/Title: Identification of the second major allergen of Japanese cedar pollen.  
A/Reference number: A60147; MUID:90342988; PMID:2382797  
A/Accession: A60147  
A/Molecule type: protein  
A/Residues: 55-64 <SAK>  
C/Keywords: glycoprotein, pollen  
F1-54/Domain: signal sequence #status predicted <SIG>  
F15-460/Product: second major allergen Cry j #status predicted <MAT>

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 81; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAYONPASWK 15  
|||:|||||  
DB 120 VDGIIAAYONPASWK 134

## RESULT 3

UC7366

Jun a 2 proteain - mountain cedar

C:Species: Juniperus ashei (mountain cedar)

C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C:Accession: UC7366; PC7093

R:Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.

Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A:Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A:Reference number: UC7366

A:Accession: UC7366

A:Molecule type: mRNA

A:Residues: 1-507 <YOK>

A:Cross-references: GB:AJ404653

A:Accession: PC7093

A:Molecule type: protein

A:Residues: 55-63 <YOK>

C:Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

o the polygalacturonase family.

C:Keywords: glycoprotein; pollen

Query Match 75.3%; Score 61; DB 2; Length 507;  
Best Local Similarity 73.3%; Pred. No. 0.012;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VDGIIAAYONPASWK 15  
|||:|||||  
DB 121 VDGIIAAYONPASWK 135

## RESULT 4

UC7100

polygalacturonase Cha o 2 - Japanese cypress

C:Species: Chamaecyparis obtusa (Japanese cypress)

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

C:Accession: UC7100; PC7026

R:Mori, T.; Yokoyama, M.; Komiyama, N.; Okano, M.; Kino, K.

Biochem. Biophys. Res. Commun. 263, 166-171, 1999

A:Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A:Reference number: UC7100; MUID:99417540; PMID:10486272

A:Accession: UC7100

A:Molecule type: mRNA

A:Residues: 1-514 <MOR>

A:Accession: PC7026

A:Molecule type: protein

A:Residues: 51-62 <MOR>

Query Match 59.3%; Score 48; DB 2; Length 514;  
Best Local Similarity 60.0%; Pred. No. 2.3;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VDGIIAAYONPASWK 15  
|||:|||||  
DB 120 VDGIIAAYONPASWK 134

## RESULT 5

T02021

hypothetical protein T9E19.2 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999

C:Accession: T02021

R:Strommatt, C.; Johnson, D.; Le, T.  
submitted to the EMBL Data Library, November 1998

A:Description: The sequence of A. thaliana T9E19.

A:Reference number: Z14496

A:Accession: T02021

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-374 <STR>

A:Cross-references: EMBL:AF104920; NID:g3859610; PID:g3859611

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 279/3

A:Note: T9E19.2

Query Match 55.6%; Score 45; DB 2; Length 374;  
Best Local Similarity 63.6%; Pred. No. 5.6;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
|||:|||||  
DB 154 VDGIIAAYONP 164

## RESULT 6

T47381

hypothetical protein T5C2.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47381

R:Obermaier, B.; Ottenwajlder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Lemcke, K.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24463

A:Accession: T47381

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1113 <OBE>

A:Cross-references: EMBL:AL138664

A:Experimental source: cultivar Columbia; BAC clone T5C2

C:Genetics:

A:Map position: 3

A:Introns: 42/1; 351/3; 418/1; 617/3; 659/3; 690/3; 732/3; 767/3; 801/3; 816/3; 845/3

A:Note: T5C2.90

Query Match 55.6%; Score 45; DB 2; Length 1113;  
Best Local Similarity 63.6%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
|||:|||||  
DB 226 VDGIIAAYONP 236

## RESULT 7

D85089

hypothetical protein AT4g08880 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C:Accession: D85089

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: AB5001; MUID:20083488; PMID:10617198

A:Accession: D85089

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1175 <STO>

A:Cross-references: GB:NC\_001268; NID:g7267530; PIDN:CAB78012.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g08880

A:Map position: 4

Query Match 55.6%; Score 45; DB 2; Length 1175;

Best Local Similarity 63.6%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
||||:|  
Db 90 VDGVAHAHNP 100

## RESULT 8

hypoetical protein F28H5.2 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C/Accession: E86402

R/Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; PMID:21016719; PMID:11130712

A/Accession: E86402

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1198 <STO>

A/Cross-references: GB:AE005172; NID:g1098930; PIDN:AAG26070.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 55.6%; Score 45; DB 2; Length 1198;  
Best Local Similarity 63.6%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
||||:|  
Db 226 VDGVAHAHNP 236

## RESULT 9

hypoetical protein F14G11.4 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C/Accession: F86386

R/Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; PMID:21016719; PMID:11130712

A/Accession: F86386

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1201 <STO>

A/Cross-references: GB:AE005172; NID:g11560181; PIDN:AAG38123.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 55.6%; Score 45; DB 2; Length 1201;  
Best Local Similarity 63.6%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
||||:|

Db 226 VDGVAHAHNP 236

## RESULT 10

hypoetical protein AT4G03300 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C/Accession: H85041

R/Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spr

Nature 402, 769-777, 1999

A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A/Reference number: A85001; PMID:20083488; PMID:10617198

A/Accession: H85041

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1285 <STO>

A/Cross-references: GB:NC\_001268; NID:g7270200; PIDN:CA877815.1; GSPDB:GN00140

C/Genetics:

A/Map position: 4

Query Match 55.6%; Score 45; DB 2; Length 1285;  
Best Local Similarity 63.6%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
||||:|  
Db 378 VDGVAHAHNP 388

## RESULT 11

hypoetical protein F5F19.8 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: H96559

R/Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; PMID:21016719; PMID:11130712

A/Accession: H96559

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1305 <STO>

A/Cross-references: GB:AE005173; NID:g4220449; PIDN:AAD12676.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 55.6%; Score 45; DB 2; Length 1305;  
Best Local Similarity 63.6%; Pred. No. 21;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYONP 11  
||||:|  
Db 226 VDGVAHAHNP 236

## RESULT 12

T47331

hypoetical protein F7P3.20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C/Accession: T47331

R/Vitale, D.; Lignori, R.; Flores, M.; Argiriou, A.; De Simone, V.; Mewes, H.W.; Rudd,

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24461  
 A:Accession: T47331  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1314 <VIT>  
 A:Cross-references: EMBL:AL138663  
 A:Experimental source: Cultivar Columbia; BAC clone F7P3  
 C:Genetics:  
 A:Map position: 3  
 A:Intons: 42/1; 351/3; 416/1; 615/3; 657/3; 712/3; 754/3; 789/3; 819/3; 839/3; 873/3;  
 A:Note: F7P3.20

Query Match 55.6%; Score 45; DB 2; Length 1314;  
 Best Local Similarity 63.6%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAYQNP 11  
 ||:|:|:|  
 Db 226 VDGIVAAHNSNP 236

## RESULT 13

T02599  
 hypothetical protein Atcg14770 [imported] - Arabidopsis thaliana  
 N:Alternate names: hypothetical protein F26C24.9  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
 C:Accession: T02599; C84521  
 R:Rounsailey, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, June 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC F26C24 genomic sequence.  
 A:Reference number: Z14660  
 A:Accession: T02599  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1756 <ROU>

A:Cross-references: EMBL:AC004705; NID:G3252804; PID:G3252818  
 A:Experimental source: Cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsailey, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Niekman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: C84521  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1756 <STO>  
 A:Cross-references: GB:AE002093; NID:G3252818; PID:AA024188.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: Atcg14770; F26C24.9  
 A:Map position: 2  
 A:Intons: 42/1; 351/3; 418/1; 617/3; 659/3; 690/3; 725/1; 758/3; 778/3; 812/3; 902/2; 9

Query Match 55.6%; Score 45; DB 2; Length 1756;  
 Best Local Similarity 63.6%; Pred. No. 28;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAYQNP 11  
 ||:|:|:|  
 Db 226 VDGIVAAHNSNP 236

## RESULT 14

T37252  
 probable matrix metalloproteinase (EC 3.4.24.-) - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
 C:Accession: T37252  
 R:Wada, K.; Sato, H.; Kinoh, H.; Kajita, M.; Yamamoto, H.; Seiki, M.  
 Gene 211, 57-62, 1998  
 A:Title: Cloning of three Caenorhabditis elegans genes potentially encoding novel matrix

A:Reference number: Z21645; MUID:98241501; PMID:9573338  
 A:Accession: T37252  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-521 <WAD>  
 A:Cross-references: EMBL:AB007817; NID:G3152405; PIDN:BA028353.1; PID:G3152406  
 A:Experimental source: strain N2  
 C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprote  
 C:Keywords: hydrolase; metalloproteinase

Query Match 54.3%; Score 44; DB 2; Length 521;  
 Best Local Similarity 40.0%; Pred. No. 12;  
 Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VDGIIAYQNPASW 15  
 ||:|:|:|  
 Db 229 IDSVMFAYDTPRCWK 243

## RESULT 15

B96971  
 serine protein kinase (prkA protein), P-loop containing [imported] - Clostridium acet  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: B96971  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Le  
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 U. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: B96971  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-646 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK78557.1; PID:G15023447; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC0579  
 A:Superfamily: Bacillus protein kinase A

Query Match 53.1%; Score 43; DB 2; Length 646;  
 Best Local Similarity 53.8%; Pred. No. 22;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 DGIILAYQNPASW 14  
 ||:|:|:|  
 Db 301 DGIILAHNSNEAW 313

Search completed: April 20, 2003, 13:15:49  
 Job time : 9.07895 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 seconds  
(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524D-97

Sequence: 1 VDGIIAAYGNPASWK 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	81	100.0	514	1	MPA2_CRYUA
2	45	55.6	373	1	ACOC_ALCEU
3	43	53.1	2073	1	BIME_EMENT
4	41	50.6	327	1	CYSM_AQUAE
5	41	50.6	561	1	PGMU_RAT
6	40.5	50.0	544	1	YR33_CABEL
7	40	49.4	230	1	TRMD_MYCTU
8	40	49.4	302	1	RE14_SCHPO
9	40	49.4	351	1	VB19_VACCD
10	40	49.4	351	1	VB19_VACCV
11	40	49.4	353	1	VB19_VACCC
12	40	49.4	444	1	GLGA_DEIRA
13	39	48.1	126	1	PRO1_DICDI
14	39	48.1	299	1	CARB_STRTH
15	39	48.1	356	1	GPDA_TREPA
16	39	48.1	561	1	PGMU_HUMAN
17	39	48.1	561	1	PGMU_MOUSE
18	39	48.1	561	1	PGMU_RABIT
19	39	48.1	3461	1	RELN_MOUSE
20	38.5	47.5	200	1	CYDX_RAT
21	38.5	47.5	327	1	YJBN_HAELN
22	38	46.9	186	1	YER7_HAELN
23	38	46.9	227	1	GYRA_SPTIT
24	38	46.9	306	1	RAM2_CANPL
25	38	46.9	440	1	SECY_CORNL
26	38	46.9	476	1	SECY_CORNL
27	38	46.9	566	1	MEHM_ECOLI
28	38	46.9	775	1	MEHM_ECOLI
29	38	46.9	775	1	MGDI_MOUSE
30	38	46.9	827	1	MGDI_MOUSE
31	38	46.9	1163	1	ITAX_YEAST
32	38	46.9	1582	1	ITAX_HUMAN
33	38	46.9	3462	1	ITAX_HUMAN

34	37	45.7	209	1	YE28_AQUAE
35	37	45.7	351	1	VB19_VARV
36	37	45.7	362	1	CNTR_CHICK
37	37	45.7	437	1	VIPS_MOUSE
38	37	45.7	593	1	US22_HUMVA
39	37	45.7	605	1	PHAC_METEX
40	37	45.7	644	1	YKAG_ECOLI
41	37	45.7	778	1	MGDI_HUMAN
42	37	45.7	821	1	GYRA_BACSU
43	37	45.7	835	1	BGAL_LYCES
44	37	45.7	1045	1	PRTS_SERMA
45	36.5	45.1	500	1	CPDG_CAVRO

## ALIGNMENTS

RESULT 1  
MPA2\_CRYUA  
ID MPA2\_CRYUA STANDARD, PRT, 514 AA.  
AC P43212;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Possible polylacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)  
DE (Major pollen allergen Cry j 2) (Cry j II).  
OS Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
OX NCBI\_Taxid=3369;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Pollen;  
RA MEDLINE=95010777; Pubmed=7926035;  
RA Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.,  
RA Usui M., Kurimoto M.;  
RT "Molecular cloning of the second major allergen, Cry j II, from  
RT Japanese cedar pollen.";  
RL FEMS Lett. 353:124-128(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pollen;  
RA MEDLINE=94271186; Pubmed=8002972;  
RA Komiyama N., Sone T., Shimizu K., Morikubo K., Kino K.;  
RT "cDNA cloning and expression of Cry j II the second major allergen of  
RT Japanese cedar pollen.";  
RL Biochem. Biophys. Res. Commun. 201:1021-1028(1994).  
RN [3]  
RP SEQUENCE OF 55-64.  
RC MEDLINE=90342988; Pubmed=2382797;  
RA Sakaguchi M., Inouye S., Taniat M., Ando S., Usui M., Matsumi T.;  
RT "Identification of the second major allergen of Japanese cedar  
RT pollen.";  
RL Allergy 45:309-312(1990).  
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-  
CC galactosiduronic linkages in pectate and other galacturonans.  
CC -1- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
CC (POLYGALACTURONASES).  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D37765; BAA07021.1; -  
CC EMBL: D29772; BAA06172.1; -  
CC HSSP: P26509; 1BHE.  
CC InterPro: IPR000743; GH28.  
CC Pfam: PF00295; Glyco\_Hydro\_28; 1.

Query March 100.0%; Score 81; DB 1; Length 514;  
Best Local Similarity 100.0%; Pred. No. 1,2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YDGI1AAAYONPASMK 15  
|||||  
DB 120 YDGI1AAAYONPASMK 134

RESULT 2  
ACOC\_ALCEU STANDARD; PRT; 373 AA.  
ID ACOC\_ALCEU STANDARD; PRT; 373 AA.  
AC P27747;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DT dihydrolipamide acetyltransferase component of acetoin cleaving  
DE system (EC 2.3.1.12) (Acetoin dehydrogenase E2 component)  
DS (Fast-migrating protein) (FMP).  
GN ACOC.  
OS Alcaligenes eutrophus (Ralstonia eutropha).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_Taxid=510;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.  
RC STRAIN=H16 / DSM 428 / ATCC 17699;  
RX MEDLINE=91286190; PubMed=2061286;  
RA Prietert H., Hein S., Knueger N., Zeh K., Schmidt B., Steinduechel A.;  
RT "Identification and molecular characterization of the Alcaligenes  
eutrophus H16 aceto operon genes involved in acetoin catabolism.",  
J. Bacteriol. 173:4056-4071(1991)  
RL  
RL 5. FUNCTION: DIHYDROLIPOAMIDE ACETYLTTRANSFERASE INVOLVED IN  
ACETOIN CATABOLISM.  
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipamide = CoA + S-  
acetyl-dihydrolipamide.  
CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL  
COFACTOR (PROBABLE).  
CC -1- PATHWAY: Acetoin catabolism.  
CC -1- INDUCTION: BY GROWTH ON ACETOIN.  
CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U66060; AAA21950.1; -.

```

DR PIR; D42462; D42462.
DR HSSP; P20708; 1GH7.
DR InterPro; IPR003089; AB_hydrolase.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR003016; Lipoyl.
DR InterPro; IPR000379; Ser_estrg_site.
DR Pfam; PF00364; biotin_lipoyl; 1.
DR Pfam; PF00561; abhydrolase; 1.
DR PRINTS; PR00111; ABHYDROLASE.
DR PROSITE; PS00189; LIPOYL; 1.
KM Transferase; Acyltransferase; Lipoyl.
FT BINDING 49 0 LIPOYL (BY SIMILARITY).
FT INT MET 0
FT BINDING 49 49
SQ SEQUENCE 373 AA; 36904 MW; 9D5CE1F83E94892F CRC64;

OY 1 VDGI1AYONPA 12
      :|::|::|:
      :|::|::|:
Db 92 IDGVAAVETPA 103

RESULT 3
ID BIEM EMENT STANDARD; PRT; 2073 AA.
AC P24686;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Negative regulator of mitosis.
GN BIME.
OS Emerticella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emerticella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90375468; PubMed=1697851;
RA Engle D.B., Osmati S.A., Osmati A.H., Rosborough S., Xiang X.,
RA Morris N.R.;
RT "A negative regulator of mitosis in Aspergillus is a putative
RT membrane-spanning protein.";
RT J. Biol. Chem. 265:16132-16137(1990).
CC -!- FUNCTION: NEGATIVE REGULATOR OF MITOSIS IN E.NIDULANS. THIS
CC PROTEIN IS PART OF A REGULATOR PATHWAY THAT INCLUDES THE NIMA
CC PROTEIN KINASE. IT IS REQUIRED TO PREVENT PREMATURE ENTRY INTO
CC MITOSIS. MUTATIONS TO THIS PROTEIN BOTH CAUSE CELLS TO ENTER
CC MITOSIS AND PREVENT THEM FROM LEAVING MITOSIS.
CC -!- SIMILARITY: TO MOUSE TSG24 AND YEAST YNL172W.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M59705; AA051478.1; -.
DR PIR; A37879; BKASBE.
DR InterPro; IPR002015; APC_proteaseome.
DR Pfam; PF01651; PC_rep; 4.
KM Transmembrane; Mitosis.
FT DOMAIN 342 353
FT TRANSMEM 1623 1643 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT TRANSMEM 1685 1703 POTENTIAL.
FT TRANSMEM 1746 1764 POTENTIAL.
SQ SEQUENCE 2073 AA; 229178 MW; 05B4E81EADDF51B4 CRC64;

Query Match 53.1%; Score 43; DB 1; Length 2073;

```

Best Local Similarity 50.0%; Pred. No. 26;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 GIYAONPASW 14  
DB 867 GVNASMNTASW 878

## RESULT 4

CYSM\_AQUAE STANDARD; PRT; 327 AA.  
ID CYSM\_AQUAE  
AC 067507;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE (O-acetylserine (Thiol)-lyase) (CSASE)  
GN CYSM OR AC1556.  
OC Aquifex aeolicus  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VFS;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.J., Keller M., Anjay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus";  
RL Nature 392:353-358(1998).  
CC -1 CATALYTIC ACTIVITY: O3-acetyl-L-serine + H(2)S = L-cysteine +  
CC acetate.  
CC -1 COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1 PATHWAY: Cysteine biosynthesis.  
CC -1 SIMILARITY: BELONGS TO THE CYSTEINE SYNTHASE/CYSTATHIONINE BETA-  
CC SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; AEO00745; AAC07459.1; -  
CC DR HSSP; P12674; 10AS.  
CC DR InterPro; IPR001926; B6\_enzyme\_beta.  
CC DR InterPro; IPR001216; Cys\_synthase.  
CC DR Pfam; PF00291; PALP; 1.  
CC DR TIGRPFAM; TIGR01136; cysKM; 1.  
CC DR TIGRPFAM; TIGR01138; cysM; 1.  
CC DR PROSITE; PS00901; CYS SYNTHASE; 1.  
CC Lysase: Cysteine biosynthesis; Pyridoxal phosphate; Complete proteome.  
CC FT BINDING 65  
CC FT SEQUENCE 327 AA; 54FC4338949FB40B CRC64;  
CC -----  
CC Query Match 50.6%; Score 41; DB 1; Length 327;  
CC Best Local Similarity 75.0%; Pred. No. 9;  
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 8 YONPASWK 15  
DB 167 YNNPASWK 174

RESULT 5  
PGMU RAT STANDARD; PRT; 561 AA.  
AC P38652;  
DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phosphoglucosyltransferase (EC 5.4.2.2) (Glucose phosphomutase) (PGM).  
GN PGM1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94040821; PubMed=8224913;  
RA Rivera A.A., Elton T.S., Dey N.B., Bonnellie P., Marchase R.B.,  
RT "Isolation and expression of a rat liver cDNA encoding  
RT phosphoglucosyltransferase";  
RL Gene 133:261-266(1993).

CC -1 FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND  
CC SYNTHESIS OF GLUCOSE.  
CC -1 CATALYTIC ACTIVITY: Alpha-D-glucose 1-phosphate = alpha-D-glucose  
CC 6-phosphate.  
CC -1 COFACTOR: Magnesium (By similarity).  
CC -1 SUBUNIT: MONOMER (By similarity).  
CC -1 SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1 SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; L11694; AAA16862.1; -  
CC DR PIR; JC2011; PMRT.  
CC DR HSSP; P00949; 3PWG.  
CC DR InterPro; IPR001485; PG/PWM\_mutase.  
CC DR Pfam; PF00408; PGM\_PWM; 1.  
CC DR Pfam; PF02878; PGM\_PWM\_1; 1.  
CC DR Pfam; PF02879; PGM\_PWM\_1; 1.  
CC DR Pfam; PF02880; PGM\_PWM\_1; 1.  
CC DR PRINTS; PR00509; PGM\_PWM.  
CC DR PROSITE; PS00710; PGM\_PWM; 1.  
CC KW Isomerase; Phosphorylation; Magnesium.  
CC FT INIT MET 0  
CC FT ACT\_SITE 116 116  
CC FT BY SIMILARITY  
CC FT FORMS THE PHOSPHOSERINE INTERMEDIATE  
CC FT SEQUENCE 561 AA; 61272 MW; 585A151D49170A66 CRC64;  
CC -----

Query Match 50.6%; Score 41; DB 1; Length 561;  
Best Local Similarity 54.5%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 IAYONPASWK 15  
DB 349 IASYEPTGWK 359

RESULT 6  
YRT3\_CAEEL STANDARD; PRT; 544 AA.  
ID YRT3\_CAEEL  
AC Q10046;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical 60.5 kDa protein T07A5.3 in chromosome III.  
GN T07A5.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2;  
 RA Buck D.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z48055; CAA88134.1; -  
 DR WormPeP; T07A5.3; CE01648.  
 KM Hypothetical protein; Transport; Symport; Sodium transport;  
 KW Transmembrane; Glycoprotein.  
 FT TRANSMEM 50 70  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 155 175 POTENTIAL.  
 FT TRANSMEM 194 214 POTENTIAL.  
 FT TRANSMEM 221 241 POTENTIAL.  
 FT TRANSMEM 286 306 POTENTIAL.  
 FT TRANSMEM 324 344 POTENTIAL.  
 FT TRANSMEM 361 381 POTENTIAL.  
 FT TRANSMEM 385 405 POTENTIAL.  
 FT TRANSMEM 417 437 POTENTIAL.  
 FT TRANSMEM 451 471 POTENTIAL.  
 SQ SEQUENCE 544 AA; 60543 MW; 4615B3C27D9F66CC CRC64;  
 Query Match 50.0%; Score 40.5; DB 1; Length 544;  
 Best Local Similarity 44.4%; Pred. No. 18;  
 Matches 8; Conservative 2; Mismatches 5; Indels 3; Gaps 1;  
 QY 1 VDGIT--AAAYNPASWK 15  
 DB 433 VGGAVTNTVTYQNDGWR 450  
 RESULT 7  
 TRMD MYCTU STANDARD; PRT; 230 AA.  
 ID TRMD\_MYCTU  
 AC Q10797;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE tRNA (Guanine-N(1)-methyltransferase (EC 2.1.1.31) (MIG-  
 DE methyltransferase) (tRNA [Gm37] methyltransferase).  
 GN TRMD OR RV2906C OR MT2974 OR MTCY274.37C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekela F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy J.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.P., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAS  
 CC (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-  
 CC homocysteine + tRNA containing N(1)-methylguanine.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z74024; CAA98343.1; -  
 DR TIGR; MT2974; -  
 DR TIGR; MT2974; -  
 DR TubercuList; RV2906c; -  
 DR InterPro; IPR002649; tRNA\_mig\_MT.  
 DR Pfam; PF01746; tRNA\_mig\_MT; 1.  
 DR ProDom; PD004978; tRNA\_mig\_MT; 1.  
 DR TIGRFAMs; TIGR00088; trmd; 1.  
 KM Transferase; Methyltransferase; tRNA processing; Complete proteome.  
 SQ SEQUENCE 230 AA; 25166 MW; 5981C3D149EDDACS CRC64;  
 Query Match 49.4%; Score 40; DB 1; Length 230;  
 Best Local Similarity 41.2%; Pred. No. 9.5;  
 Matches 7; Conservative 5; Mismatches 3; Indels 2; Gaps 1;  
 QY 1 VDGIT--AAAYNPASWK 15  
 DB 175 IDGLESPSYTRPASWR 191  
 RESULT 8  
 RE14 SCHPO STANDARD; PRT; 302 AA.  
 ID RE14\_SCHPO  
 AC Q09150;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Meiotic recombination protein rec14.  
 GN REC14 OR SPBC32F12.02.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=97403302; PubMed=9258671;  
 RA Evans D.H., Li Y.F., Fox M.E., Smith G.R.;  
 RA "A WD repeat protein, Rec14, essential for meiotic recombination in  
 RT Schizosaccharomyces pombe";  
 RL Genetics 146:1253-1264(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule L., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Glymptre B.,  
 RA Wellens I., Vanstreets E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fitch M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehach H., Reinhardt R., Pohl T.M.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Gallberg F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovskiy G.V., Useley D., Barrett B.G., Nurte P.,  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U39144; AAB71433.1; -  
 DR EMBL: AL023796; CAAL9363.1; -  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 7.  
 DR PRINTS: PR00320; GPROTEINRPT.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 7.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 3.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 DR KEGG: K01001; Repeat; WD repeat.  
 FT REPEAT 14 51 WD 1.  
 FT REPEAT 57 96 WD 2.  
 FT REPEAT 101 140 WD 3.  
 FT REPEAT 142 184 WD 4.  
 FT REPEAT 185 226 WD 5.  
 FT REPEAT 227 266 WD 6.  
 FT REPEAT 269 301 WD 7.  
 SQ SEQUENCE 302 AA; 32936 MW; 7F650E98A7E1EB7 CRC64;  
 Query Match 49.4%; Score 40; DB 1; Length 302;  
 Best Local Similarity 53.3%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 VDGIIAAYONPASWK 15  
 DB 123 VDGIIAAYONPASDK 137  
 RESULT 9  
 VB19\_VACCD STANDARD; PRT; 351 AA.  
 ID VB19\_VACCD  
 AC P23956;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Surface antigen S precursor (S antigen).  
 GN B19R.  
 OS Vaccinia virus (strain Dairen I).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 RN NCBI\_TaxID=10250;  
 RX

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90320131; PubMed=2196742;  
 RA Ueda Y., Morikawa S., Matsura Y.,  
 RT "Identification and nucleotide sequence of the gene encoding a  
 RT surface antigen induced by vaccinia virus."  
 RL Virology 177:588-594(1990).  
 CC -1- FUNCTION: MAY BIND INTERLEUKIN-1 AND/OR INTERLEUKIN-6 AND PREVENT  
 CC THESE CYTOKINES REACHING THEIR NATURAL RECEPTORS. IN CONSEQUENCE  
 CC THE INFLAMMATORY RESPONSE WOULD BE DIMINISHED AND VIRUS  
 CC REPLICATION ENHANCED.  
 CC -1- SUBCELLULAR LOCATION: INDUCED ON THE SURFACE OF VACCINIA  
 CC VIRUS-INFECTED CELLS.  
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: TO THE INTERLEUKIN-1 RECEPTOR.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: D90076; BAA1416.1; -  
 DR PIR: A35522; SAYZV.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003600; Ig\_like.  
 DR Pfam: PF00047; Ig; 2.  
 DR SMART: SM00410; Ig\_like; 1.  
 DR KEGG: K01001; Immunoglobulin domain; Antigen; Glycoprotein; Repeat; Signal;  
 KW Early protein.  
 FT SIGNAL 1 19  
 FT CHAIN 20 351  
 FT DOMAIN 65 137  
 FT DOMAIN 164 229  
 FT DOMAIN 264 341  
 FT DISULFID 73 129  
 FT DISULFID 172 221  
 FT DISULFID 221 333  
 FT CARBOHYD 117 117  
 FT CARBOHYD 182 182  
 FT CARBOHYD 261 261  
 FT CARBOHYD 269 269  
 FT CARBOHYD 321 321  
 SQ SEQUENCE 351 AA; 40701 MW; 20997CB67D39E7DB CRC64;  
 Query Match 49.4%; Score 40; DB 1; Length 351;  
 Best Local Similarity 28.6%; Pred. No. 15;  
 Matches 4; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VDGIIAAYONPASW 14  
 DB 281 IDVLIEMENPSGM 294  
 RESULT 10  
 VB19\_VACCV STANDARD; PRT; 351 AA.  
 ID VB19\_VACCV  
 AC P25213;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Surface antigen S precursor (S antigen).  
 GN B19R OR B18R.  
 OS Vaccinia virus (strain WR).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 OC NCBI\_TaxID=10254;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91259063; PubMed=2045793;  
 RX

RA Smith G.L., Chan Y.S., Howard S.T.;  
 RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near  
 RL the right inverted terminal repeat."  
 RN J. Gen. Virol. 72:1349-1376(1991).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:91170931; PubMed:1826022;  
 RA Smith G.L., Chan Y.S.;  
 RT "Two vaccinia virus proteins structurally related to the  
 RL interleukin-1 receptor and the immunoglobulin superfamily."  
 CC J. Gen. Virol. 72:511-518(1991)  
 CC -1- FUNCTION: MAY BIND INTERLEUKIN-1 AND/OR INTERLEUKIN-6 AND PREVENT  
 CC THESE CYTOKINES REACHING THEIR NATURAL RECEPTORS. IN CONSEQUENCE  
 CC THE INFLAMMATORY RESPONSE WOULD BE DIMINISHED AND VIRUS  
 CC REPLICATION ENHANCED.  
 CC -1- SUBCELLULAR LOCATION: INDUCED ON THE SURFACE OF VACCINIA  
 CC VIRUS-INFECTED CELLS.  
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: TO THE INTERLEUKIN-1 RECEPTOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D11079; BAA01848.1; -  
 CC EMBL: D01019; BAA00826.1; -  
 CC EMBL: A19579; CAA01478.1; -  
 CC PIR: B38472; SAVZWR.  
 CC PIR: JQ1812; JQ1812.  
 CC InterPro: IPR003006; IG\_MHC.  
 CC InterPro: IPR003600; IG\_Like.  
 CC Pfam: PF00047; Ig\_2.  
 CC SMART: SM00410; Ig\_Like; 1.  
 CC DR Immunoglobulin domain; Antigen; Glycoprotein; Repeat; Signal;  
 KW Early protein.  
 CC FT SIGNAL 1 19  
 CC FT CHAIN 20 351  
 CC FT DOMAIN 65 137  
 CC FT DOMAIN 164 229  
 CC FT DOMAIN 264 341  
 CC FT DISULFID 73 129  
 CC FT DISULFID 172 221  
 CC FT DISULFID 272 333  
 CC FT CARBOHYD 117 117  
 CC FT CARBOHYD 182 182  
 CC FT CARBOHYD 261 261  
 CC FT CARBOHYD 269 269  
 CC FT CARBOHYD 321 321  
 CC FT CARBOHYD 40731 MW; 73BFD6EAE1EF73F91 CRC64;  
 SQ SEQUENCE 351 AA;  
 Query Match 49.4%; Score 40; DB 1; Length 351;  
 Best Local Similarity 28.6%; Pred. No. 15;  
 Matches 4; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VDGIIAAYONPASW 14  
 DB 281 IDVLIVEMENPSGW 294  
 RESULT 11  
 VB19 VACCC STANDARD; PRT; 353 AA.  
 ID P1077;  
 AC 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Surface antigen S precursor (S antigen).  
 GN B19R. "

OS Vaccinia virus (strain Copenhagen).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 OC NCBI\_TaxId=10249;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:91021027; PubMed:2219722;  
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,  
 RA Paoletti E.;  
 RT "The complete DNA sequence of vaccinia virus."  
 RL Virology 179:247-266(1990).  
 CC [2]  
 CC COMPLETE GENOME.  
 CC RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,  
 CC Paoletti E.;  
 CC "Appendix to 'The complete DNA sequence of vaccinia virus'";  
 CC Virology 179:517-563(1990).  
 CC -1- FUNCTION: MAY BIND INTERLEUKIN-1 AND/OR INTERLEUKIN-6 AND PREVENT  
 CC THESE CYTOKINES REACHING THEIR NATURAL RECEPTORS. IN CONSEQUENCE  
 CC THE INFLAMMATORY RESPONSE WOULD BE DIMINISHED AND VIRUS  
 CC REPLICATION ENHANCED.  
 CC -1- SUBCELLULAR LOCATION: INDUCED ON THE SURFACE OF VACCINIA  
 CC VIRUS-INFECTED CELLS.  
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: TO THE INTERLEUKIN-1 RECEPTOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M35027; AAA48218.1; -  
 CC PIR: I42527; SAVZVC.  
 CC InterPro: IPR003006; IG\_MHC.  
 CC InterPro: IPR003600; IG\_Like.  
 CC Pfam: PF00047; Ig\_2.  
 CC SMART: SM00410; Ig\_Like; 1.  
 CC DR Immunoglobulin domain; Antigen; Glycoprotein; Repeat; Signal;  
 KW Early protein.  
 CC FT SIGNAL 1 21  
 CC FT CHAIN 22 353  
 CC FT DOMAIN 67 139  
 CC FT DOMAIN 166 231  
 CC FT DOMAIN 266 343  
 CC FT DISULFID 75 131  
 CC FT DISULFID 174 223  
 CC FT DISULFID 274 335  
 CC FT CARBOHYD 119 119  
 CC FT CARBOHYD 184 184  
 CC FT CARBOHYD 263 263  
 CC FT CARBOHYD 271 271  
 CC FT CARBOHYD 323 323  
 CC FT CARBOHYD 40952 MW; 51BF8080CDE1A224 CRC64;  
 SQ SEQUENCE 353 AA;  
 Query Match 49.4%; Score 40; DB 1; Length 353;  
 Best Local Similarity 28.6%; Pred. No. 15;  
 Matches 4; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VDGIIAAYONPASW 14  
 DB 283 IDVLIVEMENPSGW 296  
 RESULT 12  
 GLGA DEIRA STANDARD; PRT; 444 AA.  
 ID GLGA DEIRA  
 AC 09RWS1;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glycogen synthase (EC 2.4.1.21) (Starch [bacterial glycogen] synthase).  
 GN GLGA OR DR0594.  
 OS Deinococcus radiodurans.  
 CC Bacteria; Thermus/deinococcus group; Deinococci; Deinococcales;  
 CC Deinococcaceae; Deinococcus.  
 CX NCBI\_TaxID=1299;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffet K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Makarova K.S., Aravind L., McDonald L., Usterback T., Zalewski C.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RL radiodurans R1";  
 RL Science 286:1571-1577(1999).  
 CC -1- FUNCTION: Synthesizes alpha-1,4-glucan chains using ADP-glucose.  
 CC -1- CATALYTIC ACTIVITY: ADP-glucose + {(1,4)-alpha-D-glucosyl}(N) =  
 CC ADP + {(1,4)-alpha-D-glucosyl}(N+1).  
 CC -1- PATHWAY: Glycogen biosynthesis; second step.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001917; AAF10170.1; ALT\_INT.  
 DR TIGR; DR0594;  
 DR InterPro; IPR001296; Glycos\_transf\_1.  
 DR Pfam; PF00534; Glycos\_transf\_1; 2.  
 KW Glycogen biosynthesis; Transferase; Glycosyltransferase;  
 KW Complete proteome.  
 FT BINDING 15 ADP-GLUCOSE (BY SIMILARITY).  
 SQ SEQUENCE 444 AA; 48457 MW; 767605781A915302 CRC64;  
 Query Match 49.4%; Score 40; DB 1; Length 444;  
 Best Local Similarity 60.0%; Pred. No. 18;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 AAYONPASWK 15  
 DB 408 AAFQDPADQWQ 417  
 RESULT 13  
 PRO1 DICDI STANDARD; PRT; 126 AA.  
 AC P26199;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Profilin I.  
 GN PROA.  
 OS Dictyostelium discoideum (Slime mold).  
 CC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 CX NCBI\_TaxID=44689;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RX MEDLINE=92226170; PubMed=1725525;  
 RA Haugwitz M., Noegel A.A., Rieger D., Lottspeich F., Schleicher M.;  
 RT "Dictyostelium discoideum contains two profilin isoforms that differ

RT in structure and function".  
 RL J. Cell Sci. 100:481-489(1991).  
 CC -1- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE  
 CC CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE  
 CC POLYMERIZATION OF ACTIN. WHEREAS IT ENHANCES IT AT LOW  
 CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF  
 CC IP3 AND DG.  
 CC -1- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC  
 CC ACTIN IN A 1:1 RATIO.  
 CC -1- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X61581; CA43781.1; -.  
 DR PIR; S18027; PADO1.  
 DR HSSP; Q42449; IAOX.  
 DR DictyDB; DD01035; PROA.  
 DR InterPro; IPR002097; Profilin.  
 DR Pfam; PF00235; Profilin; 1.  
 DR SMART; SM00392; PROFILIN; 1.  
 DR PROSITE; PS00414; PROFILIN; 1.  
 KW Actin-binding; Cytoskeleton; Multigene family.  
 SQ SEQUENCE 126 AA; 13064 MW; 0A92CAD75D0CE970 CRC64;  
 Query Match 46.1%; Score 39; DB 1; Length 126;  
 Best Local Similarity 63.6%; Pred. No. 7.8;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 DGIYAONPNA 12  
 DB 44 DGIALFNPNA 54  
 RESULT 14  
 CARB\_STRTH STANDARD; PRT; 299 AA.  
 AC P13079;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE rRNA methyltransferase (EC 2.1.1.-) (Carbomycin-resistance protein).  
 GN CARB.  
 OS Streptomyces thermotolerans.  
 OG Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
 CX NCBI\_TaxID=80858;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 11416;  
 RX MEDLINE=87248111; PubMed=3036668;  
 RA Epp J.K., Burgett S.G., Schoner B.E.;  
 RT "Cloning and nucleotide sequence of a carbomycin-resistance gene from  
 RT Streptomyces thermotolerans.";  
 RL Gene 53:73-83(1987).  
 CC -1- FUNCTION: PROBABLE RNA METHYLASE. CARB CONFERS RESISTANCE TO  
 CC CARBOMYCIN AND SEVERAL OTHER MACROLIDES, LINCOMYCIN AND  
 CC VERNAMYCIN B, BUT NOT TO ALL MACROLIDE-LINCOSAMIDE-STREPTOGRAMIN  
 CC B ANTIBIOTICS.  
 CC -1- INDUCTION: THE EXPRESSION OF CARB IS INDUCIBLE BY CERTAIN  
 CC MACROLIDE ANTIBIOTICS.  
 CC -1- SIMILARITY: BELONGS TO THE RNA ADENINE N-6-METHYLTRANSFERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M16503; AAC32026.1; -  
DR PIR; A26512; A26512.  
DR InterPro; IPR001737; RNA\_A\_dimeth.  
DR InterPro; IPR000051; SAM\_Bind.  
DR Pfam; PF00398; RnaMAD; 1.  
DR PROSITE; PS01131; RNA\_A\_DIMETH; 1.  
KW Antibiotic resistance; Transferase; Methyltransferase; Plasmid.  
SQ SEQUENCE 299 AA; 34138 MW; 63990A894C044C6 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 299;  
Best Local Similarity 41.7%; Pred. No. 19;  
Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 GI1AAYQNPASW 14  
DB 278 GVVAVVTPGQW 289

RESULT 15  
GPD\_A\_TREPA STANDARD; PRT; 356 AA.

AC 083973;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-  
DE dependent glycerol-3-phosphate dehydrogenase).

GN GISA OR TP1009.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_Taxid=160;

RN 11)  
RP SEQUENCE FROM N.A.

RC STRAIN=Nichols;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren B., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,  
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;

RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete.";  
RL Science 281:375-388 (1998).

CC -1- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P) (+) =  
CC glycerone phosphate + NAD(P)H.

CC -1- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate  
CC formation.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -1- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE001269; AAC65960.1; -

DR TIGR; TP1009; -

DR InterPro; IPR001652; NAD\_Gly3P\_dh.

DR Pfam; PF01210; NAD\_Gly3P\_dh; 1.  
DR PRINTS; PR00077; GPDHDEGNASE.  
DR PRODOM; PD001649; NAD\_Gly3P\_dh; 1.  
DR PROSITE; PS00957; NAD\_G3PDH; FALSE\_NEG.

KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.

SQ SEQUENCE 356 AA; 37878 MW; 03967CE61767D915 CRC64;  
Query Match 48.1%; Score 39; DB 1; Length 356;  
Best Local Similarity 72.7%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VDGI1AAYQNP 11  
DB 163 VTGLIAASQNP 173

Search completed: April 20, 2003, 13:07:38  
Job time : 3.92105 secs



GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds  
(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524D-97

Perfect score: 81

Sequence: 1 VDGIIAAYQNPASWK 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:  
2: SP\_BACTERIA:  
3: SP\_FUNGI:  
4: SP\_HUMAN:  
5: SP\_INVERTEBRATE:  
6: SP\_MAMMAL:  
7: SP\_MHC:  
8: SP\_ORGANELLE:  
9: SP\_PHAGE:  
10: SP\_PLANT:  
11: SP RODENT:  
12: SP\_VIRUS:  
13: SP\_VERTEBRATE:  
14: SP\_UNCLASSIFIED:  
15: SP\_VIRUS:  
16: SP\_BACTERIAP:  
17: SP\_ARCHAEP:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	75.3	507	10	Q9FY19
2	45	55.6	374	10	Q9ZSH6
3	45	55.6	442	10	Q9FZP0
4	45	55.6	1113	10	Q9ML80
5	45	55.6	1119	10	Q9LKO4
6	45	55.6	1175	10	Q9ZEP3
7	45	55.6	1198	10	Q9C6N9
8	45	55.6	1201	10	Q9C607
9	45	55.6	1285	10	Q9ZRU1
10	45	55.6	1305	10	Q9ZU21
11	45	55.6	1314	10	Q9ML94
12	45	55.6	1444	10	Q9LTK8
13	45	55.6	1444	10	Q9SFT0
14	45	54.3	1756	10	Q80975
15	44	54.3	521	5	Q61266
16	43	53.1	503	10	Q941Z1

17	43	53.1	640	16	Q8XKR0	Q8XKR0 clostridium
18	43	53.1	646	16	Q97L12	Q97L12 clostridium
19	43	53.1	658	12	Q91TV8	Q91TV8 tupaia hep
20	43	53.1	1036	16	Q8YPD6	Q8YPD6 anabaena sp
21	43	53.1	1102	16	Q8YX7	Q8YX7 anabaena sp
22	43	53.1	1645	16	Q8YX3	Q8YX3 caenorhabdi
23	42	51.9	120	5	Q9TZ93	Q9TZ93 caenorhabdi
24	42	51.9	306	2	Q51333	Q51333 oesekovia x
25	42	51.9	976	3	Q96VU4	Q96VU4 blumeria gr
26	42	51.9	1140	16	Q8U964	Q8U964 agrobacteri
27	42	51.9	2628	16	Q8V0G6	Q8V0G6 yellow head
28	41	50.6	352	12	Q8VX9	Q8VX9 anabaena sp
29	41	50.6	352	12	Q8VX2	Q8VX2 monkeypox v
30	41	50.6	355	10	Q94KI6	Q94KI6 pisum sativ
31	41	50.6	355	12	Q8V2H8	Q8V2H8 camelox vi
32	41	50.6	465	10	Q9ASN3	Q9ASN3 oryza sativ
33	41	50.6	465	16	Q8Z7B0	Q8Z7B0 salmonella
34	41	50.6	467	16	Q8ZP63	Q8ZP63 salmonella
35	40	49.4	94	16	Q8X868	Q8X868 escherichia
36	40	49.4	155	5	Q27190	Q27190 tetrahymena
37	40	49.4	200	16	Q9WY15	Q9WY15 thermotoga
38	40	49.4	212	5	Q9TXB7	Q9TXB7 drosophila
39	40	49.4	249	16	Q55394	Q55394 synecocyst
40	40	49.4	315	12	Q91FV7	Q91FV7 bean 58058
41	40	49.4	343	16	Q9RC51	Q9RC51 bacillus ha
42	40	49.4	351	12	Q72754	Q72754 cowpox viru
43	40	49.4	353	12	Q9UF34	Q9UF34 vaccinia vi
44	40	49.4	358	12	Q9UFS5	Q9UFS5 ectromella
45	40	49.4	358	12	Q9LBP4	Q9LBP4 ectromella

## ALIGNMENTS

RESULT 1	ID	Q9FY19	PRELIMINARY;	PRT;	507 AA.
AC	Q9FY19	01-MAR-2001 (TREMUR1.16, Created)			
DT	01-MAR-2001 (TREMUR1.16, Last sequence update)				
DT	01-JUN-2002 (TREMUR1.21, Last annotation update)				
DE	Pollen major allergen 2 protein precursor.				
GN	JNA2.				
OS	Juniperus ashei (Ozark white cedar).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.				
OX	NCBI_TaxID=13101;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=MALE POLLEN;				
RX	MEDLINE=20403896; PubMed=10944464;				
RA	Yokoyama M., Miyahara M., Shimizu K., Kino K., Tsunoo H.;				
RT	Purification, identification and cDNA cloning of Jun a 2, the second				
RT	major allergen of mountain cedar pollen."				
RL	Biochem. Biophys. Res. Commun. 275:195-202(2000).				
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES				
CC	(POLYGALACTURONASES)				
DR	EMBL: A404653; CAC05582.1; -				
DR	HSSP: P26509; IBBE.				
DR	InterPro: IPR00743; GH28.				
DR	InterPro: IPR000408; Reg. chr. condens.				
DR	Pfam: PF00295; Glyco. hydro. 28; 1.				
DR	PROSITE: PS00502; POLYGALACTURONASE; UNKNOWN_1.				
KW	PROSITE: PS00626; RC1_2; UNKNOWN_1.				
KW	Cell wall; Glycosidase; Hydrolase; Signal.				
FT	SIGNAL				
FT	1				
FT	POTENTIAL.				
FT	54				
FT	SEQUENCE				
FT	507 AA; 55730 MW; 2B2E0A5B58F5A CRC64;				
FT	Query Match				
FT	Best Local Similarity				
FT	Matches				
FT	11; Conservative				
FT	1; Mismatches				
FT	3; Indels				
FT	0; Gaps				
FT	0;				

Db 121 VDGITAAVDPKAKWK 135

## RESULT 2

Q9ZSH6 PRELIMINARY; PRT; 374 AA.  
AC Q9ZSH6; 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DE T9E19.2 protein.  
GN T9E19.2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Washu;  
RT "The A. thaliana Genome Sequencing Project."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Stromatt C., Johnson D., Le T.;  
RT "The sequence of A. thaliana T9E19."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Waterston R.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF104920; AAC72876.1; -. 479ECD55DFA7A953 CRC64;  
SQ SEQUENCE 374 AA; 42429 MW; 479ECD55DFA7A953 CRC64;

Query Match 55.6%; Score 45; DB 10; Length 374;  
Best Local Similarity 63.6%; Pred. No. 14;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDGITAAVONP 11  
Db 154 VDGVAHAHNP 164

## RESULT 3

Q9FZP0 PRELIMINARY; PRT; 442 AA.  
AC Q9FZP0; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE Emb|CAB81580.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB028613; BAB10789.1; -. 8568236E28808E8A CRC64;  
SQ SEQUENCE 442 AA; 50198 MW; 8568236E28808E8A CRC64;

Query Match 55.6%; Score 45; DB 10; Length 442;  
Best Local Similarity 63.6%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDGITAAVONP 11  
Db 226 VDGVAHAHNP 236

## RESULT 4

Q9M180 PRELIMINARY; PRT; 1113 AA.  
AC Q9M180; 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE Hypothetical 123.6 kDa protein.  
GN 15C2\_90.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Obermaier B., Ottenwaelde B., Duchemin D., Zeitler K., Mewes H.W.,  
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL138664; CAB81580.1; -.  
DR InterPro; IPR001304; Lectin C.  
DR PROSITE; PS00615; C TYPE LECTIN\_1; UNKNOWN\_1.  
KM Hypothetical protein.  
SQ SEQUENCE 1113 AA; 123579 MW; B826A0588B97DDB CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1113;  
Best Local Similarity 63.6%; Pred. No. 51;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDGITAAVONP 11  
Db 226 VDGVAHAHNP 236

## RESULT 5

Q9LK04 PRELIMINARY; PRT; 1119 AA.  
AC Q9LK04; 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE Gb|AAC24188.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA MEDLINE-20363099; PubMed-10907853;  
RX Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
Sequence features of the regions of 4,251,695 bp covered by ninety pl.  
TAC and BAC clones."  
RL DNA Res. 7:217-221 (2000).  
DR EMBL; AP000382; BAB02940.1; -.  
DR InterPro; IPR003653; SUMO\_protease.

DR Pfam; PF02902; Peptidase\_C48; 1.  
 SQ SEQUENCE 1119 AA; 125821 MW; CD3483A918B0895 CRC64;  
 Query Match 55.6%; Score 45; DB 10; Length 1119;  
 Best Local Similarity 63.6%; Pred. No. 51;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAYONP 11  
 |||:|:|:  
 Db 226 VDGVAHNSNP 236

RESULT 6  
 ID Q9ZPF3 PRELIMINARY; PRT; 1175 AA.  
 AC Q9ZPF3;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE T3H13.8 protein.  
 GN T3H13.8 OR A74G08880.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA WASHU;  
 RT "The A. thaliana Genome Sequencing Project.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Drone K.; Nguyen C.;  
 RT "The sequence of A. thaliana T3H13.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Waterston R.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A7128396; AAD17370.1;  
 DR EMBL; AL161513; CAB78012.1;  
 DR InterPro; IPR003653; SUMO\_protease.  
 DR Pfam; PF02902; Peptidase\_C48; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1175 AA; 131492 MW; AD1924338690E242 CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1175;  
 Best Local Similarity 63.6%; Pred. No. 54;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAYONP 11  
 |||:|:|:  
 Db 90 VDGVAHNSNP 100

RESULT 7  
 ID Q9C6N9 PRELIMINARY; PRT; 1198 AA.  
 AC Q9C6N9;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 133.6 kDa protein.  
 GN F2815.2.  
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A.; Ecker J.R.; Palm C.J.; Federspiel N.A.; Kaul S.;  
 RA White O.; Alonso J.; Altati H.; Araujo R.; Bowman C.L.; Brooks S.Y.;  
 RA Buehler E.; Chan A.; Chao Q.; Chen H.; Cheuk R.F.; Chin C.W.;  
 RA Chung M.K.; Conn L.; Conway A.B.; Conway A.R.; Creasy T.H.; Dewar K.;  
 RA Dunn P.; Egu P.; Feldblum T.V.; Feng J.-D.; Fong B.; Fujii C.Y.;  
 RA Gill J.E.; Goldsmith A.D.; Haas B.; Hansen N.F.; Hughes B.; Hutzar L.;  
 RA Hunter J.L.; Jenkins J.; Johnson-Hopson C.; Khan S.; Khaykin E.;  
 RA Kim C.J.; Koo H.L.; Kremenetskaia I.; Kurtz D.B.; Kwan A.; Lam B.;  
 RA Langin-Hooper S.; Lee A.; Lee J.M.; Lenz C.A.; Li J.H.; Li Y.-P.;  
 RA Lin X.; Liu S.X.; Liu Z.A.; Lueros J.S.; Maiti R.; Marziani A.;  
 RA Miltescher J.; Miranda M.; Nguyen M.; Nierman W.C.; Osborne B.T.;  
 RA Pai G.; Peterson J.; Pham P.K.; Rizzo M.; Rooney T.; Rowley D.;  
 RA Sakano H.; Salzberg S.L.; Schwartz J.R.; Shinn P.; Southwick A.M.;  
 RA Sun H.; Tallon L.J.; Tambunga G.; Toriumi M.J.; Town C.D.;  
 RA Utterback T.; Van Aken S.; Vaysberg M.; Vysotskaya V.S.; Walker M.;  
 RA Wu D.; Yu G.; Fraser C.M.; Venter J.C.; Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";  
 RL Nature 408:816-820 (2000).  
 DR EMBL; AC079280; AAG50582.1;  
 DR InterPro; IPR003653; SUMO\_protease.  
 DR Pfam; PF02902; Peptidase\_C48; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1198 AA; 133579 MW; D608AE577CE21339 CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1198;  
 Best Local Similarity 63.6%; Pred. No. 55;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAYONP 11  
 |||:|:|:  
 Db 226 VDGVAHNSNP 236

RESULT 8  
 ID Q9C607 PRELIMINARY; PRT; 1201 AA.  
 AC Q9C607;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 134.0 kDa protein.  
 GN F14G11.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A.; Ecker J.R.; Palm C.J.; Federspiel N.A.; Kaul S.;  
 RA White O.; Alonso J.; Altati H.; Araujo R.; Bowman C.L.; Brooks S.Y.;  
 RA Buehler E.; Chan A.; Chao Q.; Chen H.; Cheuk R.F.; Chin C.W.;  
 RA Chung M.K.; Conn L.; Conway A.B.; Conway A.R.; Creasy T.H.; Dewar K.;  
 RA Dunn P.; Egu P.; Feldblum T.V.; Feng J.-D.; Fong B.; Fujii C.Y.;  
 RA Gill J.E.; Goldsmith A.D.; Haas B.; Hansen N.F.; Hughes B.; Hutzar L.;  
 RA Hunter J.L.; Jenkins J.; Johnson-Hopson C.; Khan S.; Khaykin E.;  
 RA Kim C.J.; Koo H.L.; Kremenetskaia I.; Kurtz D.B.; Kwan A.; Lam B.;  
 RA Langin-Hooper S.; Lee A.; Lee J.M.; Lenz C.A.; Li J.H.; Li Y.-P.;  
 RA Lin X.; Liu S.X.; Liu Z.A.; Lueros J.S.; Maiti R.; Marziani A.;  
 RA Miltescher J.; Miranda M.; Nguyen M.; Nierman W.C.; Osborne B.T.;  
 RA Pai G.; Peterson J.; Pham P.K.; Rizzo M.; Rooney T.; Rowley D.;  
 RA Sakano H.; Salzberg S.L.; Schwartz J.R.; Shinn P.; Southwick A.M.;

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Uteback T., Van Aken S., Vayberg M., Vysotskaya V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana"; 816-820 (2000).  
 RL Nature 408:816-820 (2000).  
 DR EMBL; AC084221; AAG50521.1; -;  
 DR InterPro; IPR003653; SUMO\_protease.  
 DR Pfam; PF02902; Peptidase\_C48; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1201 AA; 133995 MW; D10P9847289CFB7 CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1201;  
 Best Local Similarity 63.6%; Pred. No. 56;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYQNP 11  
 |||:|:|  
 DB 226 VDGVAHAHNP 236

## RESULT 9

Q9ZR01 PRELIMINARY; PRT; 1285 AA.

AC Q9ZR01 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 143.8 kDa protein.  
 GN F4C21.23 OR AT4G03300.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucots II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Habermann K., de la Bastide M., Huang E.N., Gnoj L., Schurz K.,  
 RA Preston R., Calma C., Martienssen R., Parnell L.D., Dedhia N.,  
 RA McCombie W.R.;  
 RT "Arabidopsis thaliana BAC F4C21 from chromosome IV near 17 cM";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,  
 RA Preston R.R., Mateo A., Shah R., O'Shaughnessy A., Rodriguez M.,  
 RA Sheher M., Schurz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005275; AAD14458.1; -;  
 DR EMBL; AL161496; CAB77815.1; -;  
 DR InterPro; IPR003653; SUMO\_protease.  
 DR Pfam; PF02902; Peptidase\_C48; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1285 AA; 143833 MW; 7D1542C2E7E0F5DC CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1285;  
 Best Local Similarity 63.6%; Pred. No. 60;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYQNP 11  
 |||:|:|  
 DB 378 VDGVAHAHNP 388

RESULT 10  
 Q9ZU21 PRELIMINARY; PRT; 1305 AA.  
 AC Q9ZU21;

DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE F5F19.8 protein.  
 GN F5F19.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucots II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Vysotskaya V.S., Schwartz J.R., Toriumi M., Yu G., Lenz C., Liu S.,  
 RA Li J., Kremetskaia I., Iuros J., Gonzalez A., Altafi H., Araujo R.,  
 RA Buehler B., Conway A.B., Conn L., Dunn P., Hansen N., Huizar L.,  
 RA Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,  
 RA Ecker J.R., Federle N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F5F19 sequence";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006216; AAD12676.1; -;  
 DR InterPro; IPR003653; SUMO\_protease.  
 DR Pfam; PF02902; Peptidase\_C48; 1.  
 SQ SEQUENCE 1305 AA; 146080 MW; D3C036AEE2EFD6C CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1305;  
 Best Local Similarity 63.6%; Pred. No. 61;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYQNP 11  
 |||:|:|  
 DB 226 VDGVAHAHNP 236

RESULT 11  
 Q9M194 PRELIMINARY; PRT; 1314 AA.

AC Q9M194 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 146.5 kDa protein.  
 GN F7P3\_20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucots II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Vitale D., Ligouri R., Flores M., Argirou A., De Simone V.,  
 RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,  
 RA Salanoubat M.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL138663; CAB86693.1; -;  
 DR InterPro; IPR003653; SUMO\_protease.  
 DR Pfam; PF02902; Peptidase\_C48; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1314 AA; 146506 MW; 534C779C9A11EE CRC64;

Query Match 55.6%; Score 45; DB 10; Length 1314;  
 Best Local Similarity 63.6%; Pred. No. 62;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDGIIAAYQNP 11  
 |||:~|:~|  
 DB 226 VDGVAHAHNP 236

RESULT 12

```

09LUG8
ID Q9LUG8      PRELIMINARY;      PRT; 1444 AA.
AC Q9LUG8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similarity to mutator-like transposase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025605; BAA98062.1; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
SQ SEQUENCE 1444 AA; 161024 MW; AF0CB23AAAC4EB8B CRC64;

Query Match
Best Local Similarity 55.6%; Score 45; DB 10; Length 1444;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

CY 1 VDGIIAYONP 11
DB 226 VDGVAASHNP 236

RESULT 13
Q9SFY0
ID Q9SFY0      PRELIMINARY;      PRT; 1444 AA.
AC Q9SFY0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE T22C5.24.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA Shiao P., Alafati H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharly N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Tortumt M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T22C5 from chromosome
RT 1."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC012375; AAF24952.1; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
SQ SEQUENCE 1444 AA; 161065 MW; C0563453E81C6498 CRC64;

Query Match
Best Local Similarity 55.6%; Score 45; DB 10; Length 1444;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

CY 1 VDGIIAYONP 11
DB 226 VDGVAASHNP 236

RESULT 14

```

```

080975
ID 080975      PRELIMINARY;      PRT; 1756 AA.
AC 080975;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE At2g14770 protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Bentro M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.U., Rongming C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gili J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Coppenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004705; AAC24188.1; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
SQ SEQUENCE 1756 AA; 197368 MW; 68FC3B5FCC282775 CRC64;

Query Match
Best Local Similarity 55.6%; Score 45; DB 10; Length 1756;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

CY 1 VDGIIAYONP 11
DB 226 VDGVAASHNP 236

RESULT 15
061266
ID 061266      PRELIMINARY;      PRT; 521 AA.
AC 061266;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Matrix metalloproteinase.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=98241501; PubMed=9573338;
RA Wada K., Sato H., Kinoh H., Kajita M., Yamamoto H., Seki M.;
RT "Cloning of three Caenorhabditis elegans genes potentially encoding
RT novel matrix metalloproteinases."
RL Gene 211:57-62 (1998)
DR EMBL; AB007817; BAA28353.1; -
DR HSRF; P03956; ICGU.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrxin.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase_M10; 1.

```

Sat May 10 08:44:51 2003

us-09-142-524d-97.rpt

Page 6

DR SMART; SM00120; HX; 4.  
DR SMART; SM00235; ZINC; 1.  
DR PROSITE; PS00546; CYSTEINE\_SWITCH; UNKNOWN\_1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 521 AA; 60700 MW; F3936A1C732C6FE2 CRC64;

Query Match 54.3%; Score 44; DB 5; Length 521;  
Best Local Similarity 40.0%; Pred. No. 31;  
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VDGIIAAYQNPASWK 15  
:|::|||  
Db 229 IDSVWFAYDTPRKWK 243

Search completed: April 20, 2003, 13:13:10  
Job time : 15.6711 secs

GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07695 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524d-100

Perfect score: 82

Sequence: 1 NNRIWLOPAKLTGFT 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73.\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*  
5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	82	100.0	514	2	S48730	Cry j II protein - second major aller
2	82	100.0	514	2	JC2498	polygalacturonase
3	58	70.7	514	2	JC7100	Jun a 2 protein -
4	52	63.4	507	2	JC7366	protein F28C11.9 (
5	47	57.3	1161	2	B86368	probable polygalac
6	43	52.4	466	2	H96728	hypothetical prote
7	42	51.2	466	2	A12283	hypothetical prote
8	42	51.2	492	2	C96521	protein F21D18.18
9	42	51.2	540	2	B96631	probable polygalac
10	42	51.2	853	2	S59315	hypothetical prote
11	41	50.0	63	2	D34514	MHC class II histo
12	41	50.0	235	2	D65071	hypothetical prote
13	41	50.0	235	2	B91097	hypothetical prote
14	41	50.0	235	2	A85943	hypothetical prote
15	41	50.0	452	1	D69810	phosphotransferase
16	40	48.8	246	2	F72464	hypothetical prote
17	40	48.8	756	2	S74742	hypothetical prote
18	39	47.6	217	2	AB3432	exopolysaccharide
19	39	47.6	225	2	T39083	thiol:disulfide in
20	39	47.6	495	2	G82371	conserved hypotnet
21	39	47.6	677	2	T26574	Flxg-related prote
22	39	47.6	880	2	G90163	hypothetical prote
23	39	47.6	1037	2	B65022	DNA-directed RNA p
24	39	47.6	1037	2	D91045	acriiflavin resist
25	39	47.6	1037	2	G85889	antimycoside eff
26	39	47.6	1037	2	AE0816	hypothetical prote
27	38	46.3	328	2	T24705	probable efflux pu
28	38	46.3	328	2	G64610	hypothetical prote
29	38	46.3	328	2	G71904	probable transcrip

30	38	46.3	420	2	C71407	hypothetical prote
31	38	46.3	479	2	UQ0781	sucrose uptake pro
32	38	46.3	487	1	QOECRS	ylgE protein - Bac
33	38	46.3	549	2	F82415	conserved hypotnet
34	38	46.3	624	2	S50650	hypothetical prote
35	38	46.3	790	2	T50337	homolog to drosoph
36	38	46.3	1044	2	B72338	galactosyltransfer
37	37.5	45.7	493	2	G69468	carbohydrate kinas
38	37.5	45.7	525	2	H81386	ABC transporter At
39	37.5	45.7	1035	2	AD3203	autotransporter pr
40	37	45.1	131	2	AE0126	conserved hypotnet
41	37	45.1	174	2	B72748	hypothetical prote
42	37	45.1	184	1	G64018	hypothetical prote
43	37	45.1	219	2	F71363	probable amino aci
44	37	45.1	241	2	T27636	hypothetical prote
45	37	45.1	299	2	F97277	probable glycosylt

#### ALIGNMENTS

##### RESULT 1

S48730

Cry j II protein - Japanese cedar

C:Species: Cryptomeria japonica (Japanese cedar)

C:Date: 07-May-1995 #sequence\_rev150n 21-Jul-1995 #text\_change 28-May-1999

C:Accession: S48730

R:Namba, M.; Kurose, M.; Torijoe, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.;

FEBS Lett. 353, 124-128, 1994

A:Title: Molecular cloning of the second major allergen, Cry j II, from Japanese ceda

A:Reference number: S48730; PMID:95010777; PMID:7926035

A:Accession: S48730

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-514 <NAM>

A:Cross-references: GB:D37765; NID:9577695; PIDN:BAA07021.1; PID:d1007598; PID:g57769

Query Match 100.0%; Score 82; DB 2; Length 514;

Best Local Similarity 100.0%; Pred.No. 1.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNRIWLOPAKLTGFT 15

Db 135 NNRIWLOPAKLTGFT 149

##### RESULT 2

JC2498

second major allergen Cry j II precursor - Japanese cedar

C:Species: Cryptomeria japonica (Japanese cedar)

C:Date: 16-Mar-1995 #sequence\_rev150n 26-May-1995 #text\_change 21-Jul-2000

C:Accession: JC2498; PC2346; A60147

R:Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994

A:Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanes

A:Reference number: JC2498; PMID:94271186; PMID:8002972

A:Accession: JC2498

A:Molecule type: mRNA

A:Residues: 1-514 <KOM>

A:Cross-references: DDBJ:D29772; NID:9506857; PIDN:BAA06172.1; PID:g506858

A:Accession: PC2346

A:Molecule type: protein

A:Residues: 52-61 <KO2>

R:Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matuhasi, T.

Allergy 45, 309-312, 1990

A:Title: Identification of the second major allergen of Japanese cedar pollen.

A:Reference number: A60147; PMID:9034298; PMID:2382797

A:Accession: A60147

A:Molecule type: protein

A:Residues: 55-64 <SAK>

C:Keywords: glycoprotein; pollen

F1-54/Domain: signal sequence #status predicted <SIG>

F155-460/Product: second major allergen Cry j #status predicted <MNT>

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 82; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNRIMLOPAKLTGFT 15  
|||||  
DB 135 NNRIMLOPAKLTGFT 149

## RESULT 3

JC7100 polygalacturonase Cha o 2 - Japanese cypress

C:Species: Chamaecyparis obtusa (Japanese cypress)  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

A:Accession: JC7100; PMID:10486272

A:Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A:Reference number: JC7100; PMID:10486272

A:Accession: JC7100

A:Molecule type: mRNA

A:Residues: 1-514 <MOR>

A:Accession: PC7026

A:Molecule type: protein

A:Residues: 51-62 <MO2>

Query Match 70.7%; Score 58; DB 2; Length 514;  
Best Local Similarity 71.4%; Pred. No. 0.022;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNRIMLOPAKLTGFT 14  
|||||  
DB 135 NNRIMLOPAKLTGFT 148

## RESULT 4

JC7366 Jun a 2 protein - mountain cedar

C:Species: Juniperus ashei (mountain cedar)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

A:Accession: JC7366; PMID:10486272

A:Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A:Reference number: JC7366

A:Accession: JC7366

A:Molecule type: mRNA

A:Residues: 1-507 <YOK>

A:Accession: PC7093

A:Molecule type: protein

A:Residues: 55-63 <YOK>

C:Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

C:Keywords: glycoprotein; pollen

Query Match 63.4%; Score 52; DB 2; Length 507;  
Best Local Similarity 57.1%; Pred. No. 0.26;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNRIMLOPAKLTGFT 14  
|||||  
DB 136 NNRIMLOPAKLTGFT 149

## RESULT 5

B86368 protein P28C11.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: B86368

R:theologus, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, T.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marzida,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: B86368

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 11161 <STO>

A:Cross-references: GB:AE005172; NID:98778576; PIDN:AAF79584.1; GSPDB:GN00141

C:Gene: P28C11.9

A:Map position: 1

Query Match 57.3%; Score 47; DB 2; Length 1161;  
Best Local Similarity 72.7%; Pred. No. 4.9;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 RIMLOPAKLTG 13  
|||||  
DB 824 RIMLOPAKLTG 834

## RESULT 6

H96728 probable polygalacturonase P24J13.7 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

A:Accession: H96728

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: H96728

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-468 <STO>

A:Cross-references: GB:AE005173; NID:96175142; PIDN:AAF04869.1; GSPDB:GN00141

C:Gene: P24J13.7

A:Map position: 1

Query Match 52.4%; Score 43; DB 2; Length 468;  
Best Local Similarity 72.7%; Pred. No. 9.7;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 RIMLOPAKLTG 13  
|||||  
DB 147 RIMLOPAKLTG 157

## RESULT 7

AI2283 hypothetical protein alr3824 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AI2283

R:kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriy

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata



DNA Res. 8, 205-213, 2001  
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: A12283  
 A/Status: Preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-266 <KOR>  
 A/Cross-references: GB:BA000019; PIDN:BA075523.1; PID:G17132958; GSPDB:GN00179  
 A/Experimental source: strain FCC 7120  
 C/Genetics:  
 A/Gene: alr3824

Query Match  
 Best Local Similarity 51.2%; Score 42; DB 2; Length 266;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 NR1WLOFAXLGTG 13  
 DB 206 NR1WLRWFTLTG 217

RESULT 8  
 C96521  
 Protein F21D18.18 (imported) - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C/Accession: C96521  
 R/Theologian: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A86141; MUID:21016719; PMID:11130712  
 A/Accession: C96521  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-492 <STO>  
 A/Cross-references: GB:AE005173; NID:98778527; PIDN:AAF79535.1; GSPDB:GN00141  
 C/Genetics:  
 A/Gene: F21D18.18  
 A/Map position: 1

Query Match  
 Best Local Similarity 51.2%; Score 42; DB 2; Length 492;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 NNR1WLOFAXLGTGFT 15  
 DB 159 NKNQWLVFVRLDGT 173

RESULT 9  
 B96631  
 Probable polygalacturonase P8A5.12 (imported) - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C/Accession: B96631  
 R/Theologian: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: B96631  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-540 <STO>  
 A/Cross-references: GB:AE005173; NID:92462753; PIDN:AA071972.1; GSPDB:GN00141  
 C/Genetics:  
 A/Gene: P8A5.12  
 A/Map position: 1

Query Match  
 Best Local Similarity 51.2%; Score 42; DB 2; Length 540;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 WLOFAXLGTGFT 15  
 DB 205 WIFFTLSGIT 215

RESULT 10  
 S59315  
 Hypothetical protein YLR127c - yeast (*Saccharomyces cerevisiae*)  
 N/Alternate names: hypothetical protein L3108  
 C/Species: *Saccharomyces cerevisiae*  
 C/Date: 29-Nov-1995 #sequence\_revision 23-Feb-1996 #text\_change 29-Oct-1999  
 C/Accession: S59315; S64964; S64969; S69414  
 R/Delius, H.  
 Submitted to the EMBL Data Library, June 1995  
 A/Description: 36.8 kb of *S.cerevisiae* chromosome XII including ACE2, CK11, PDC5, SLG1  
 A/Reference number: S59315  
 A/Accession: S59315  
 A/Molecule type: DNA  
 A/Residues: 1-853 <DEU>  
 A/Cross-references: EMBL:X91258; NID:99956686; PIDN:CAA62638.1; PID:G995689

A/Experimental source: strain S288C  
 R/Verhasselt, P.; Voet, M.; Volckaert, G.  
 Submitted to the Protein Sequence Database, May 1996  
 A/Reference number: S64964  
 A/Accession: S64964  
 A/Molecule type: DNA  
 A/Residues: 1-853 <VER>  
 A/Cross-references: EMBL:X73299; NID:91360534; PIDN:CAA97696.1; PID:e245812; PID:g1360

A/Experimental source: strain S288C  
 R/Delius, H.; Hebling, U.  
 Submitted to the Protein Sequence Database, May 1996  
 A/Reference number: S64967  
 A/Accession: S64969  
 A/Molecule type: DNA

A/Residues: 1-853 <DEM>  
 A/Cross-references: EMBL:X73299; NID:91360534; PIDN:CAA97696.1; PID:e245812; PID:g1360  
 A/Experimental source: strain S288C  
 R/Verhasselt, P.; Volckaert, G.  
 Submitted to the EMBL Data Library, September 1995  
 A/Reference number: S69393  
 A/Accession: S69414  
 A/Molecule type: DNA  
 A/Residues: 1-853 <VEN>  
 A/Cross-references: EMBL:X89514; NID:91297019; PIDN:CAA61705.1; PID:e198828; PID:g1297

C/Genetics:  
 A/Gene: SGD:RS11  
 A/Cross-references: SGD:S0004117; MIPS:YLR127C  
 A/Map position: 12R

Query Match  
 Best Local Similarity 51.2%; Score 42; DB 2; Length 853;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NR1WLOFAXLGTG 13  
 DB 248 NQYMSQFSXLVG 259

RESULT 11  
 D34514

MHC class II histocompatibility antigen, I - rhesus macaque (fragment)  
 C/Species: Macaca mulatta (rhesus macaque)  
 C/Date: 22-Jun-1990 #sequence\_revision 22-Jun-1990 #text\_change 12-Jul-1996  
 C/Accession: D34514  
 R/Gyllenstein, U.B.; Erlich, H.A.  
 A/Title: Ancient roots for polymorphism at the HLA-DQA1 locus in primates.  
 A/Reference number: A34512; MUID:9009387; PMID:2513578  
 A/Accession: D34514  
 A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A/Molecule type: DNA  
 A/Residues: 1-63 <GVL>  
 C/Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 41; DB 2; Length 63;  
 Best Local Similarity 54.5%; Pred. No. 2.6;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WLOPAKLTGFT 14  
 Db 30 LWEPKRGGF 40

RESULT 12  
 D65071  
 hypothetical protein b2876 - Escherichia coli (strain K-12)  
 C/Species: Escherichia coli  
 C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C/Accession: D65071  
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; MUID:97426617; PMID:9278503  
 A/Accession: D65071  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-235 <BLAT>  
 A/Cross-references: GB:AE00371; GB:U00096; NID:q1789239; PIDN:AACT5914.1; PID:91789241;  
 A/Experimental source: strain K-12, substrain MG1655  
 C/Superfamily: Escherichia coli hypothetical protein b2876

Query Match 50.0%; Score 41; DB 2; Length 235;  
 Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 WLOPAKLTGFT 15  
 Db 147 WSOFAADITGFT 157

RESULT 13  
 E91097  
 hypothetical protein ECG3749 [imported] - Escherichia coli (strain O157:H7, substrain RI  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C/Accession: E91097  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Accession: E91097  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-235 <HAY>  
 A/Cross-references: GB:BA000007; PIDN:BA037172.1; PID:G13363221; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RIMD 0509952  
 C/Genetics:  
 A/Gene: ECG3749  
 C/Superfamily: Escherichia coli hypothetical protein b2876

Query Match 50.0%; Score 41; DB 2; Length 235;

Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 WLOPAKLTGFT 15  
 Db 147 WSOFAADITGFT 157

RESULT 14  
 A85943  
 hypothetical protein Z4215 [imported] - Escherichia coli (strain O157:H7, substrain EL  
 C/Species: Escherichia coli  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: A85943  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh  
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: A85943  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-235 <STO>  
 A/Cross-references: GB:AE005174; NID:G12517397; PIDN:AAG58005.1; GSPDB:GN00145; UWGP:  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: Z4215  
 C/Superfamily: Escherichia coli hypothetical protein b2876

Query Match 50.0%; Score 41; DB 2; Length 235;  
 Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 WLOPAKLTGFT 15  
 Db 147 WSOFAADITGFT 157

RESULT 15  
 D69810  
 phosphotransferase system enzyme II (EC 2.7.1.69) factor II homolog yJlf - Bacillus su  
 C/Species: Bacillus subtilis  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C/Accession: D69810  
 R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bery  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, F  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iulio, M  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin  
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl  
 A/Authors: Schleich, S.; Schreier, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akeruch, M.; Tanakoshi, A.; Tanaka, T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchiya  
 T.; Witters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A/Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: D69810  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-452 <KUN>  
 A/Cross-references: GB:Z99108; GB:AL009126; NID:G2633055; PIDN:CAB12599.1; PID:G263305  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Gene: yJlf  
 C/Function:  
 A/Description: mediates transport of glucose across the cytoplasmic membrane concomit  
 ion  
 C/Superfamily: phosphotransferase system glucose-specific enzyme II, factor II; phosph  
 C/Keywords: phosphoprotein; phosphotransferase; sugar transport system; transmembrane

F;1452/Domain: phosphotransferase system glucose-specific enzyme II, factor II homology  
F;12-28/Domain: transmembrane #status predicted <TM01>  
F;46-62/Domain: transmembrane #status predicted <TM02>  
F;94-110/Domain: transmembrane #status predicted <TM03>  
F;133-149/Domain: transmembrane #status predicted <TM04>  
F;253-269/Domain: transmembrane #status predicted <TM05>  
F;286-302/Domain: transmembrane #status predicted <TM06>  
F;329-345/Domain: transmembrane #status predicted <TM07>  
F;419-436/Domain: transmembrane #status predicted <TM08>  
F;397/Active site: Cys (phosphocysteine intermediate) #status predicted

## Query Match

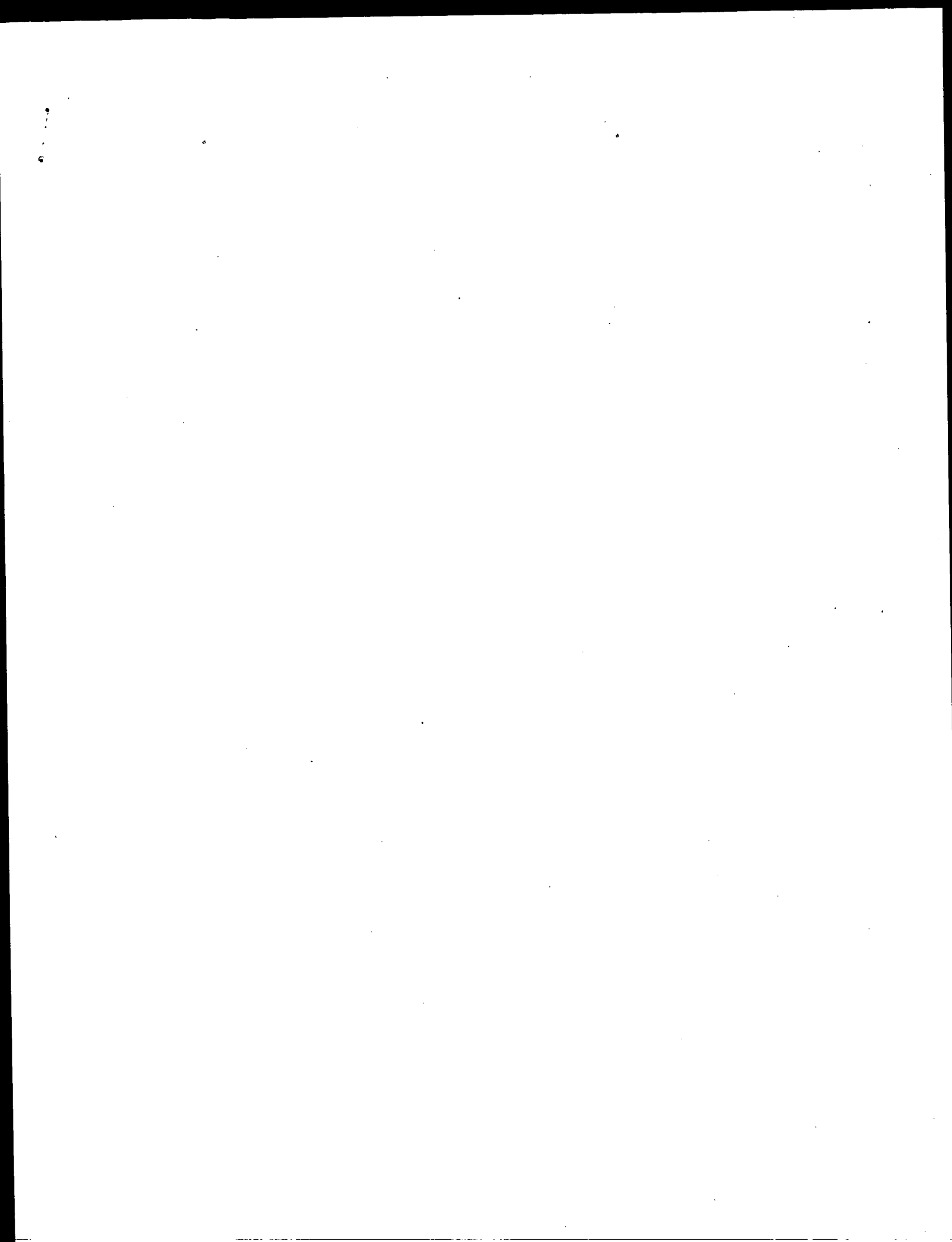
Best Local Similarity 50.0%; Score 41; DB 1; Length 452;  
Pred. No. 21;

Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 NNRIMLPKLTGFT 15

Db 192 NNIFWFOFGFYNGVT 206

Search completed: April 20, 2003, 13:15:51  
Job time : 8.07895 secs



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524D-100

Perfect score: 82

Sequence: 1 NNR1WLOPAKLTGFT 15

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	514	1	MPA2_CRYVA
2	41	50.0	256	1	YOEC_ECOLI
3	39	47.6	880	1	RBAL_SULSO
4	39	47.6	1037	1	ACRD_ECOLI
5	38	46.3	479	1	P24177 escherichia
6	38	46.3	487	1	PTSB_VIBAL
7	38	46.3	624	1	YX71_YEAST
8	37	45.1	172	1	NADM_AERPE
9	37	45.1	181	1	YGGT_HABIN
10	37	45.1	310	1	PP5_PSRSE
11	37	45.1	359	1	OTC_EWENT
12	37	45.1	361	1	OTC_ASPT
13	37	45.1	370	1	OTC_ASPT
14	37	45.1	409	1	HMDH_SULSO
15	37	45.1	490	1	EXP7_STRN
16	37	45.1	778	1	CALR_HUMAN
17	37	45.1	982	1	ENV_VILV
18	37	45.1	983	1	ENV_VILV
19	37	45.1	989	1	ENV_VILV
20	37	45.1	990	1	ENV_VILV
21	37	45.1	991	1	ENV_VILV
22	37	45.1	1353	1	CYA9_HUMAN
23	37	45.1	1353	1	CYA9_HUMAN
24	36	44.5	372	1	PROB_THETN
25	36	43.9	122	1	SAA3_RABIT
26	36	43.9	178	1	RLIX_CASSA
27	36	43.9	178	1	RLIX_CASSA
28	36	43.9	178	1	RLIX_CASSA
29	36	43.9	255	1	UNG_HSV2
30	36	43.9	315	1	CPP2_ENTHI
31	36	43.9	322	1	GM2_SCHPO
32	36	43.9	354	1	YHL4_YEAST
33	36	43.9	368	1	HIS8_BUCAL

## ALIGNMENTS

RESULT 1	MPA2_CRYVA	STANDARD:	PRT:	514 AA.
AC	P43212;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DE	Possible polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)			
DE	(Major pollen allergen Cry j 2) (Cry j II).			
OS	Cryptomeria japonica (Japanese cedar).			
OC	Burkholderia japonica (Japanese cedar).			
OC	Spermatophyta; Coniferales; Cupressaceae; Cryptomeria.			
OX	NCBI_TaxID=3369;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Pollen;			
RX	MEDLINE=9501077; PubMed=7926035;			
RA	Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.,			
RT	Usui M., Kurimoto M.;			
RT	"Molecular cloning of the second major allergen, Cry j II, from			
RT	Japanese cedar pollen.";			
RL	FEBS Lett. 353:124-128 (1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pollen;			
RX	MEDLINE=94271186; PubMed=8002972;			
RA	Komiyama N., Sone T., Shimizu K., Morikubo K., Kino K.;			
RT	"CDNA cloning and expression of Cry j II the second major allergen of			
RT	Japanese cedar pollen.";			
RL	Biochem. Biophys. Res. Commun. 201:1021-1028 (1994).			
RN	[3]			
RP	SEQUENCE OF 55-64.			
RC	Sakaguchi M., Inouye S., Tanai M., Ando S., Usui M., Matubasi T.;			
RT	"Identification of the second major allergen of Japanese cedar			
RT	pollen.";			
RL	Allergy 45:309-312 (1990).			
CC	-I- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-			
CC	galactosiduronic linkages in pectate and other galacturonans.			
CC	-I- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).			
CC	-I- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES			
CC	(POLYGALACTURONASES).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; D37765; BAA07021.1; -			
DR	EMBL; D29772; BAA06172.1; -			
DR	HSRP; P26509; IHR.			
DR	InterPro: IPR000743; GH28.			
DR	Pfam: PF00295; Glyco_hydro_28; 1.			

DR PROSITE; PS00502; POLYGLACTURONASE; 1.  
 KW Hydrolase; Glycosidase; Cell wall; Signal; Zymogen; Fruit ripening;  
 KW Amyloplast; Glycoprotein; Allergen.  
 FT SIGNAL 1 22  
 FT PROPEP 23 45  
 FT CHAIN 46 433  
 FT PROPEP 434 514  
 FT ACT SITE 278 278  
 FT CARBOHYD 460 460  
 FT CARBOHYD 472 472  
 FT CONFLICT 5 5  
 FT CONFLICT 12 12  
 FT CONFLICT 34 35  
 FT CONFLICT 37 37  
 FT CONFLICT 88 88  
 FT CONFLICT 98 98  
 FT CONFLICT 451 451  
 FT CONFLICT 454 454  
 FT CONFLICT 504 504  
 FT CONFLICT 507 507  
 SQ SEQUENCE 514 AA; 56645 MW; 624611C3FAD6302 CRC64;

Query Match 100.0%; Score 82; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NNRWLOFALKTGFT 15  
 DB 135 NNRWLOFALKTGFT 149

RESULT 2  
 YOEBC\_ECOLI STANDARD; PRT; 256 AA.  
 AC 046809;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypochemical protein ygec.  
 OS YOEBC OR B2876.  
 OC Escherichia coli.  
 OS Escherichia coli.  
 OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U28375; AAA83057.1; ALT\_INIT.  
 DR EMBL; AB000371; AAC75914.1; ALT\_INIT.  
 DR EcoGene; EG13059; ygec.  
 KW Hypochemical protein; Complete proteome.  
 SQ SEQUENCE 256 AA; 28103 MW; F9E5C04EB26ED68B CRC64;

Query Match 50.0%; Score 41; DB 1; Length 256;  
 Best Local Similarity 63.6%; Pred. No. 6.3;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 WLOFALKTGFT 15  
 DB 168 WSGFADITGFT 178

RESULT 3  
 RPAI\_SULSO STANDARD; PRT; 880 AA.  
 AC Q980R2;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6).  
 GN RPOA1 OR SSO0225.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RA MEDLINE=21332296; PubMed=11427726;  
 RA Sme Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweyer M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Erasuo G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -1- COFACTOR: ZINC (By similarity).  
 CC -1- SIMILARITY: THE COMBINED A'+A" SUBUNITS CORRESPOND TO THE A  
 CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE  
 CC EUBACTERIAL BETA' SUBUNIT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB006559; AAK40566.1; -  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Zinc-finger; Complete proteome.  
 FT ZN FING 58 101  
 SQ SEQUENCE 880 AA; 99674 MW; 5C8AE20EFC71DE25 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 880;  
 Best Local Similarity 46.2%; Pred. No. 52;  
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 3 RYWLQFALKTGFT 15  
 DB 633 RYRFRVLEQFT 645

RESULT 4  
 ACRD\_ECOLI STANDARD; PRT; 1037 AA.  
 AC P24177; P77178; Q46715; P76971;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable aminoglycoside efflux pump (Acriflavine resistance protein  
D).  
GN ACRD OR B2470.  
OS Escherichia coli.  
OC Bacterioides; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Nilles M.L., Bertrand K.P.;  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Ma D., Cook D.N., Alberti M., Nikaido H., Hearst J.E.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RA MEDLINE=97426617; PubMed=9278503;  
RA Blotner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474 (1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horikuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features.";  
RL DNA Res. 4:91-113 (1997).  
RN [5]  
RP SEQUENCE OF 998-1037 FROM N.A.  
RC STRAIN=K12;  
RA MEDLINE=92355499; PubMed=1644752;  
RA Bouvier J., Richard C., Higgins W., Bogler O., Stragier S.;  
RT "Cloning, characterization, and expression of the *dapB* gene of  
RT Escherichia coli.";  
RL J. Bacteriol. 174:5265-5271 (1992).  
RN [6]  
RP FUNCTION.  
RA MEDLINE=20158894; PubMed=10692383;  
RA Rosenberg E.Y., Ma D., Nikaido H.;  
RT "AcrD of Escherichia coli is an aminoglycoside efflux pump.";  
RL J. Bacteriol. 182:1754-1756 (2000).  
CC -1- FUNCTION: PARTICIPATES IN THE EFFLUX OF AMINOGLYCOSIDES. CONFERS  
CC RESISTANCE TO A VARIETY OF THESE SUBSTANCES.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC EMBL; U12598; AAA20584.1; -;  
CC EMBL; U10436; AAA74741.1; -;  
CC EMBL; AE000334; AAC75523.1; -;  
CC EMBL; D90875; BAA16344.1; -;  
CC EMBL; D90876; BAA16348.1; -;  
CC DR

DR EMBL; X57403; CAA40663.1; -;  
DR PIR; C42959; C42959.  
DR PIR; S26997; S26997.  
DR EcoGene; EG10014; acrd.  
DR InterPro; IPR001036; Acriflavin\_res.  
DR InterPro; IPR004764; HAE1.  
DR Pfam; PF00873; ACR\_tran\_1.  
DR PRINTS; PRO0702; ACRIFLAVINRP.  
DR TIGRFAMs; TIGR00915; 2A0602; 1.  
KW Transmembrane; Inner membrane; Transport; Complete proteome.  
FT DOMAIN 1 9  
FT TRANSSEM 10 28  
FT DOMAIN 29 339  
FT TRANSSEM 340 359  
FT DOMAIN 360 365  
FT TRANSSEM 366 385  
FT DOMAIN 386 391  
FT TRANSSEM 392 413  
FT DOMAIN 414 441  
FT TRANSSEM 442 460  
FT DOMAIN 461 473  
FT TRANSSEM 474 496  
FT DOMAIN 497 537  
FT TRANSSEM 538 556  
FT DOMAIN 557 870  
FT TRANSSEM 871 890  
FT DOMAIN 891 896  
FT TRANSSEM 897 916  
FT DOMAIN 917 922  
FT TRANSSEM 923 944  
FT DOMAIN 945 971  
FT TRANSSEM 972 990  
FT DOMAIN 991 1003  
FT TRANSSEM 1004 1026  
FT DOMAIN 1027 1037  
FT CONFLICT 303 303  
FT CONFLICT 372 372  
FT CONFLICT 385 385  
FT CONFLICT 461 461  
FT CONFLICT 665 665  
FT CONFLICT 763 763  
FT CONFLICT 775 775  
FT CONFLICT 778 778  
SQ SEQUENCE 1037 AA; 113047 MW; 961611E1D24FD45 CRC64;  
  
Query Match 47.6%; Score 39; DB 1; Length 1037;  
Best Local Similarity 66.7%; Pred. No. 62;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 3 RIWLOFAKLTF 14  
DB 185 RIWLDPAKLNKF 196  
  
RESULT 5  
PTSB\_VIBAL STANDARD; PRT; 479 AA.  
AC P22825;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE PTS system, sucrose-specific IIBC component (EIIIC-SCR) (Sucrose-  
DE permease IIBC component) (Phosphotransferase enzyme II, BC component)  
DE (EC 2.7.1.69) (EII-SCR).  
GN SCRA.  
OS Vibrio alginolyticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_Taxid=663;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=91071601; PubMed=2174811;  
RA Blatch G.L., Schollie R.R., Woods D.R.;  
RT "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose

RT uptake-encoding region." ;  
 RL Gene 95:17-23(1990).  
 RN [2]  
 RP SEQUENCE OF 1-12 FROM N.A.  
 RX MEDLINE=91285433; PubMed=2060795;  
 RA Blatch G.L., Woods D.R.;  
 RT "Nucleotide sequence and analysis of the Vibrio alginolyticus ser  
 repressor-encoding gene (serR)." ;  
 RL Gene 101:45-50(1991).  
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 TRANSPORT SYSTEM. THE IIID DOMAINS CONTAIN THE SUGAR BINDING SITE  
 AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY  
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIB TRANSFERS ITS  
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 THE SUGAR.  
 CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
 histidine + sugar phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1- SIMILARITY: CONTAINS 1 PTS EIIB DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M76768; AAA27555.1; -  
 DR EMBL; M35009; AAA27557.2; -  
 DR PIR; JQ0781; JQ0781.  
 DR HSSP; P05053; IIBA.  
 DR InterPro; IPR004719; PTS\_EIIC\_glc.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR TIGRFAMS; TIGR00826; EIIB; 1.  
 DR TIGRFAMS; TIGR00852; pts-glc; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KM Phosphotransferase system: Sugar transport; Transferase;  
 Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 ? EIIB DOMAIN.  
 FT MOD RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 324 324 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 479 AA; 49890 MW; FBF906B5170E3EB7 CRC64;  
 Query Match 46.3%; Score 38; DB 1; Length 479;  
 Best Local Similarity 42.9%; Pred. No. 41;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Oy 2 NR1WLOFAKLTGFT 15  
 Db 71 NOVYAEAKLTGMS 84  
 RESULT 6  
 TDTT ECOLI STANDARD; PRT; 487 AA.  
 AC P39414; O46870;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative taurate carrier (Taurate transporter) (Taurate/succinate  
 antiporter).  
 GN YGUE OR B3063.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subphylum; Enterobacteriaceae;  
 OC Escherichia.

OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Berra N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12." ;  
 RL Science 277:1453-1474(1997).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RP MEDLINE=87248073; PubMed=3297921;  
 RX Nesin M., Lupski J.R., Svec P., Godson G.N.;  
 RT "Possible new genes as revealed by molecular analysis of a 5-kb  
 Escherichia coli chromosomal region 5' to the rpsU-dnaG-rpoD  
 macromolecular-synthesis operon." ;  
 RL Gene 51:149-161(1987).  
 CC [3]  
 CC IDENTIFICATION.  
 RP MEDLINE=95075659; PubMed=7984428;  
 RX Borodovsky M., Rudd K.E., Koonin E.V.;  
 RA "Intrinsic and extrinsic approaches for detecting genes in a  
 bacterial genome." ;  
 RT Nucleic Acids Res. 22:4756-4767(1994).  
 RN [4]  
 RP POSSIBLE FUNCTION.  
 RX MEDLINE=98361905; PubMed=9696764;  
 RA Fos K.M., Dimroth P., Bott M.;  
 RT "The Escherichia coli citrate carrier CitT: a member of a novel  
 eubacterial transporter family related to the 2-oxoglutarate/malate  
 translocator from spinach chloroplasts." ;  
 RL J. Bacteriol. 180:4160-4165(1998).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE UPTAKE OF TARTRATE IN EXCHANGE TO  
 THE EFFLUX OF SUCCINATE (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 (probable).  
 CC -1- SIMILARITY: BELONGS TO THE NAD(P)H DEHYDROGENASE FAMILY OF TRANSPORTERS.  
 CC SODIUM SUBFAMILY.  
 CC -1- CAUTION: REP. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 IN POSITIONS 35, 51, 132, 245, 268 AND 443.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U28379; AAA89143.1; -  
 DR EMBL; AE000388; AAC76099.1; -  
 DR EMBL; M16194; -; NOT ANNOTATED\_CDS.  
 DR EcoGene; EG12393; YGJE.  
 DR InterPro; IPR001898; Na/sulF\_symp.  
 DR Pfam; PF00938; Na\_sulph\_symp; 1.  
 DR TIGRFAMS; TIGR00785; dass; 1.  
 KM Hypothetical protein; Transmembrane; Inner membrane; Transport;  
 Antiport; Complete proteome.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.  
 FT TRANSMEM 54 74 POTENTIAL.  
 FT TRANSMEM 93 113 POTENTIAL.  
 FT TRANSMEM 137 157 POTENTIAL.  
 FT TRANSMEM 189 209 POTENTIAL.  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.  
 FT TRANSMEM 313 333 POTENTIAL.  
 FT TRANSMEM 340 360 POTENTIAL.  
 FT TRANSMEM 370 390 POTENTIAL.  
 FT TRANSMEM 393 413 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.



FT TRANSMEM 465 485 POTENTIAL.  
 FT CONFLICT 404 404 L -> P (IN REF. 2).  
 FT CONFLICT 457 457 A -> T (IN REF. 2).  
 SQ SEQUENCE 487 AA; 52906 MW; 4D6E8439BEE281D2 CRC64;

Query Match  
 Best Local Similarity 46.3%; Score 38; DB 1; Length 487;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 NR1WLOPAKLTG 13  
 DB 31 NHTWLYFAVFTG 42

RESULT 7  
 TEXT YEAST STANDARD; PRT; 624 AA.

AC P40090;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 72.1 kDa protein in FTR1-SPT15 intergenic region.  
 GN YER147C.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
 RA Aviles E., Banno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,  
 RA Hyman R., Kayser A., Kump C., Laskari D., Lew H., Lin D.,  
 RA Meesele D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yelton M., Borstein D., Davis R.W.,  
 RA Submitted (DEC-1994) to the EMBL/Genbank/DBD databases.  
 RL -1- SIMILARITY: BELONGS TO THE EMBL/Genbank/DBD databases.  
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL; U18917; AAB64674.1; -.  
 DR SGD; S0000949; YER147C.  
 DR InterPro; IPR000873; AMP-bind.  
 DR PROSITE; PS00455; AMP\_BINDING; 1.  
 KW Hypothetical protein; Ligase.  
 SQ SEQUENCE 624 AA; 72140 MW; 1269D931C656E608 CRC64;

Query Match  
 Best Local Similarity 46.3%; Score 38; DB 1; Length 624;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 NR1WLOPAKLTG 11  
 DB 461 NKVWLOVEKL 470

RESULT 8  
 NADM AERPE STANDARD; PRT; 172 AA.

AC Q9YER8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nicotinamide-nucleotide adenylyltransferase (EC 2.7.7.1) (NAD(+))

DE Pyrophosphorylase (NAD(+)) diphosphorylase (NMN adenylyltransferase).  
 GN APO513.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;  
 OC Desulfurococcaceae; Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=K1.  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Akai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1."  
 RL DNA Res. 6:83-101(1999).  
 CC -1- CATALYTIC ACTIVITY: ATP + nicotinamide ribonucleotide =  
 CC diphosphate + NAD(+).  
 CC -1- PATHWAY: NAD BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL NMN ADENYLYLTRANSFERASE  
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; APO00059; BAA79478.1; ALT\_INIT.  
 DR HSSP; Q57961; 1F9A.  
 DR InterPro; IPR004821; Cyt tran rel.  
 DR InterPro; IPR004820; Cytidylyltransf.  
 DR Pfam; PF01467; Cytidylyltransf; 1.  
 DR TIGRPFAM; TIGR00125; Cyt tran rel; 1.  
 KW Transferase; Nucleotidyltransferase; NAD; Complete proteome.  
 SQ SEQUENCE 172 AA; 19828 MW; E100C434A2064E4A CRC64;

Query Match  
 Best Local Similarity 45.1%; Score 37; DB 1; Length 172;  
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 NR1WLOPAKLTG 11  
 DB 82 NKVWVQYLKM 91

RESULT 9  
 YGCT HAEIN STANDARD; PRT; 181 AA.

AC P44097;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein H11036.  
 GN H11036.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```

RA Weidmann J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO E. COLI YGGT.
CC -1- SIMILARITY: BELONGS TO THE YGGT FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U32784; AAC22696.1; ALT_INIT.
CC TIGR: H11036;
DR InterPro: IPR003425; unk_yggT.
DR Pfam: PF03235; YGGT; 2. _YGGT.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
SQ SEQUENCE 181 AA; 20200 MW; 3F086A7FEA972232 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 181;
Best Local Similarity 66.7%; Pred. No. 2;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0

OY 3 RIWLOFAKL 11
DB 24 RAWLOFAKRV 32

RESULT 10
PBPS_PSEAE
ID PBPS_PSEAE STANDARD; PRT; 310 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Penicillin-binding protein 5 precursor (PBP-5) [D-alanyl-D-alanine-
DE endopeptidase] (EC 3.4.99.-) (DD-endopeptidase).
DE PBP3 OR PA0869.
GN PBP3 OR PA0869.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.T., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: CELL WALL FORMATION.

```

```
-- -- --  
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).  
CC  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL; U62582; AAC46341.1; -.  
DR AEMBL; AB004521; AAA04258.1; -.  
DR MEROPS; S11.002;- .  
DR InterPro; IPR001967; Ala/AlacBptasel.  
DR Pfam; PF00768; Peptidase_S11_1.  
DR PRINTS; PR00725; DMDACBPASL.  
KW Hydroxylase; Peptidoglycan synthesis; Cell wall; Signal; Periplasmic;  
KM Complete proteome.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 310 PENICILLIN-BINDING PROTEIN 5.  
FT ACT SITE 66 66 ACYLATED BY PENICILLIN (BY SIMILARITY).  
FT CONFLICT 296 296 D -> DD (IN REF. 1).  
SQ SEQUENCE 310 AA; 34046 MW; B1F6ZA3A2FE9EF641 CRC64;  
  
Query Match 45.1%; Score 37; DB 1; Length 310;  
Best Local Similarity 56.2%; Pred. NO. 38;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 2;  
  
QY 1 NNRIV-LQFAKLTGFT 15  
||| : | |||  
Db 220 NKKTNIQLTK-TGFT 234  
  
RESULT 11  
OTC_EMENT ID STANDARD; PRT; 359 AA.  
AC OT_C EMCNT  
AD PI1803;  
DT 01-OCT-1989 (Rel. 12, Created)  
DI 01-OCT-1989 (Rel. 12, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DS Ornithine carbamoyltransferase, mitochondrial precursor (EC 2.1.3.3)  
DE (OnCase) (Ornithine transcarbamylase).  
GN ARGX  
OS Escherichia nidulans (Aspergillus nidulans).  
OC Bakerycota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emmentella.  
OX NCBI_TaxId=5072;  
RN [1]_  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87014107; PubMed=3020372;  
RA Upshall A., Gilbert T., Saari G., O'Hara P.J., Weglenski P., Berse B.,  
RT Miller K., Timberlake W.E.;  
RL "Molecular analysis of the argX gene of Aspergillus nidulans";  
MOL. Genet. 204:349-354(1986).  
CL CATALYTIC ACTIVITY: Carbamoyl phosphate + L-orntihline = phosphate  
+ L-citrulline.  
CC -1- PATHWAY: Arginine biosynthesis.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -1- SIMILARITY: BELONGS TO THE ATCSSES/OTCASES FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
DR EMBL; M29819; AAA50816.1; -.  
DR PIR; S07317; OMASN.
```

DR HSSP; P00480; 10TH.  
 DR InterPro; IPR002029; Asp/Orn\_Cotranf.  
 DR InterPro; IPR002292; Orn\_carbtransf.  
 DR Pfam; PF00185; OTCace; 1.  
 DR Pfam; PF02729; OTCace\_N; 1.  
 DR PRINTS; PR00100; AOTCase.  
 DR TIGRfams; TIGR00658; orn\_carb\_tr; 1.  
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.  
 DR TRANSFERASE; Arginine biosynthesis; Mitochondrion; Transit peptide.  
 FT TRANSIT 1 24  
 FT CHAIN 25 359 ORNITHINE CARBAMOYLTRANSFERASE.  
 SQ SEQUENCE 359 AA; 39058 MW; 6B2CBD7BC276F2C4 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 359;  
 Best Local Similarity 46.2%; Pred. No. 45;  
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 NR1WQFAKLTGF 14  
 Db 339 NRLMAAISALEGF 351

## RESULT 12

OTC\_ASPTG STANDARD; PRT; 361 AA.  
 AC 000291;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ornithine carbamoyltransferase, mitochondrial precursor (EC 2.1.3.3) (OTCase) (Ornithine transcarbamylase).  
 GN ARG-1 OR ARG1.  
 OS Aspergillus terreus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=31178;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97286514; PubMed=9141661;  
 RT Ventura L., Perez-Gonzalez J.A., Ramon D.;  
 RT "Cloning and molecular analysis of the Aspergillus terreus arg1 gene coding for an ornithine carbamoyltransferase";  
 RT FEMS Microbiol. Lett. 149:207-212(1997).  
 CC -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine = phosphate + L-citrulline.  
 CC -1- PATHWAY: Arginine biosynthesis.  
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; Z67741; CAA91554.1; -.  
 DR HSSP; P00480; 10TH.  
 DR InterPro; IPR002029; Asp/Orn\_Cotranf.  
 DR InterPro; IPR002292; Orn\_carbtransf.  
 DR Pfam; PF00185; OTCace; 1.  
 DR Pfam; PF02729; OTCace\_N; 1.  
 DR PRINTS; PR00100; AOTCase.  
 DR TIGRfams; TIGR00658; orn\_carb\_tr; 1.  
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.  
 DR TRANSFERASE; Arginine biosynthesis; Mitochondrion; Transit peptide.  
 FT TRANSIT 1 24  
 FT CHAIN 25 361 ORNITHINE CARBAMOYLTRANSFERASE.  
 SQ SEQUENCE 361 AA; 39129 MW; 9306AD75E499D75 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 361;

Best Local Similarity 46.2%; Pred. No. 45;  
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 NR1WQFAKLTGF 14  
 Db 341 NRLMAAISALEGF 353

## RESULT 13

OTC\_ASPTG STANDARD; PRT; 370 AA.  
 AC P11066;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ornithine carbamoyltransferase, mitochondrial precursor (EC 2.1.3.3) (OTCase) (Ornithine transcarbamylase).  
 GN ARG1.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88167829; PubMed=3443301;  
 RT Buxton F.P., Gwynne D.I., Garven S., Sibley S., Davies R.W.;  
 RT "Cloning and molecular analysis of the ornithine carbamoyl transferase gene of Aspergillus niger";  
 RT Gene 60:255-266(1987).  
 CC -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine = phosphate + L-citrulline.  
 CC -1- PATHWAY: Arginine biosynthesis.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M19158; AAA32688.1; -.  
 DR PIR; A27362; OMASG.  
 DR HSSP; P00480; 10TH.  
 DR InterPro; IPR002029; Asp/Orn\_Cotranf.  
 DR InterPro; IPR002292; Orn\_carbtransf.  
 DR Pfam; PF00185; OTCace; 1.  
 DR Pfam; PF02729; OTCace\_N; 1.  
 DR PRINTS; PR00100; AOTCase.  
 DR TIGRfams; TIGR00658; orn\_carb\_tr; 1.  
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.  
 DR TRANSFERASE; Arginine biosynthesis; Mitochondrion; Transit peptide.  
 FT TRANSIT 1 38  
 FT CHAIN 39 370 ORNITHINE CARBAMOYLTRANSFERASE.  
 SQ SEQUENCE 370 AA; 39924 MW; 77A8A8B28F6A65F CRC64;

Query Match 45.1%; Score 37; DB 1; Length 370;  
 Best Local Similarity 46.2%; Pred. No. 46;  
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 NR1WQFAKLTGF 14  
 Db 350 NRLMAAISALEGF 362

## RESULT 14

HMDH\_SULSO STANDARD; PRT; 409 AA.  
 ID HMDH\_SULSO  
 AC 008424; G9UWT6;  
 DT 01-NOV-1997 (Rel. 35, Created)



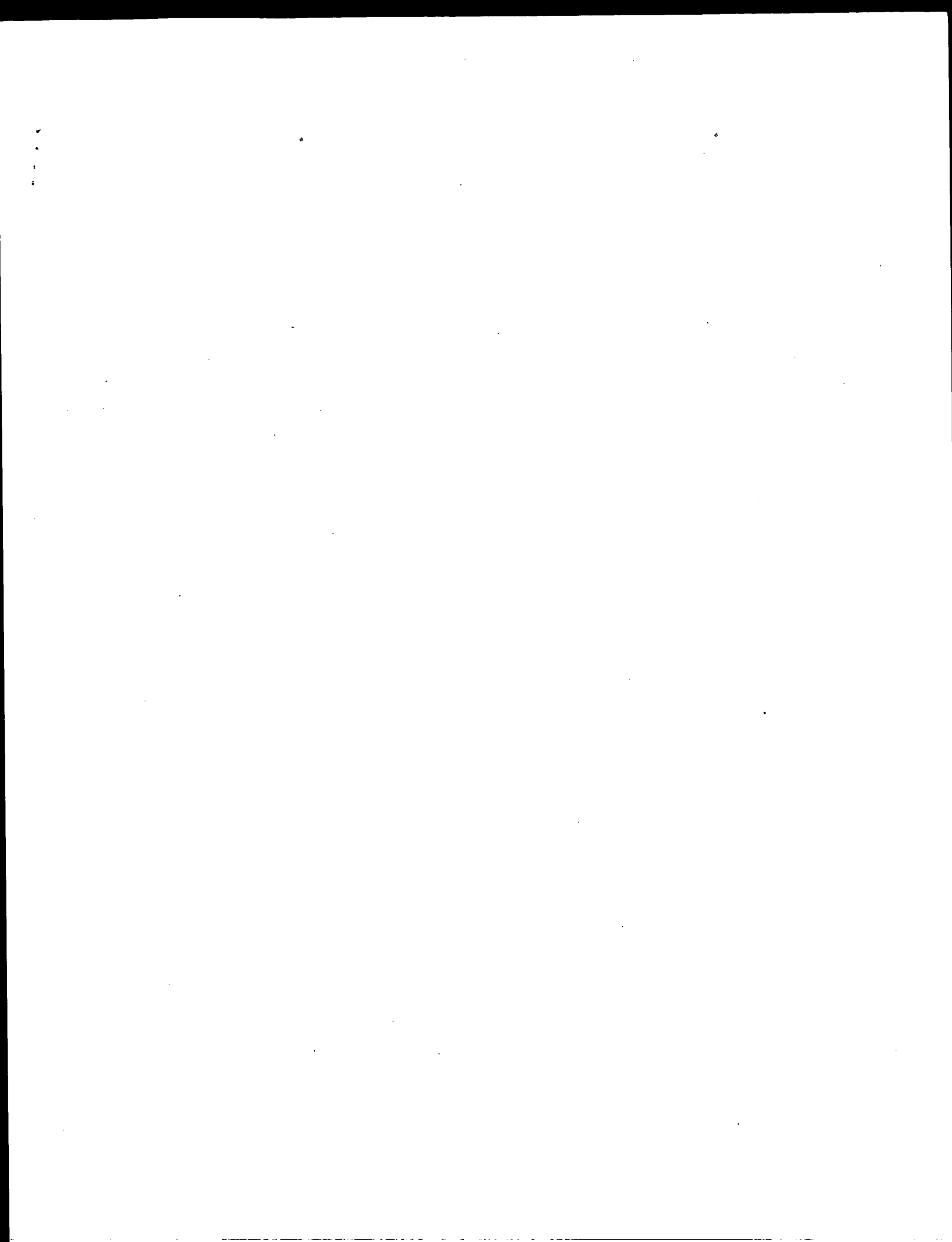
RX PubMed-10570950;  
 RA Nishikawa T., Ishikawa H., Yamamoto S., Koshihara Y.;  
 RT "A novel calcitonin receptor gene in human osteoclasts from normal  
 RT bone marrow.";  
 RL FEBS Lett. 458:409-414(1999).  
 RN [7]  
 RP ALTERNATIVE SPLICING.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95237133; PubMed=7720653;  
 RA Nussezeveig D.R., Mathew S., Gershengorn M.C.;  
 RT "Alternative splicing of a 48-nucleotide exon generates two isoforms  
 RT of the human calcitonin receptor.";  
 RL Endocrinology 136:2047-2051(1995).  
 RN [8]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=95251678; PubMed=773946;  
 RA Nakamura M., Hashimoto T., Nakajima T., Ichi S., Furuyama J.,  
 RT Ishihara Y., Kakudo K.;  
 RT "A new type of human calcitonin receptor isoform generated by  
 RT alternative splicing.";  
 RL Biochem. Biophys. Res. Commun. 209:744-751(1995).  
 RN [9]  
 RP VARIANT LEU-463.  
 RX MEDLINE=99036684; PubMed=9817931;  
 RA Taboulet J., Frenkian M., Frendo J.L., Feingold N., Jullienne A.,  
 RT de Vernejoul M.C.;  
 RT "Calcitonin receptor polymorphism is associated with a decreased  
 RT fracture risk in post-menopausal women.";  
 RL Hum. Mol. Genet. 7:2129-2133(1998).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR CALCITONIN. THE ACTIVITY OF  
 CC CYCLASE. THE CALCITONIN RECEPTOR IS THOUGHT TO COUPLE TO THE  
 CC HETEROTRIMERIC GUANOSINE TRIPHOSPHATE-BINDING PROTEIN THAT IS  
 CC SENSITIVE TO CHOLERA TOXIN.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L00587; AAA35640.1; -;  
 DR EMBL; X69920; CAA49541.1; -;  
 DR EMBL; U26553; AAC50300.1; -;  
 DR EMBL; U26554; AAC50301.1; -;  
 DR EMBL; X82466; CAA57849.1; -;  
 DR EMBL; ACC03078; AAB83944.1; -;  
 DR EMBL; ACC03078; AAB83945.1; -;  
 DR EMBL; S77876; AAD14268.1; -;  
 DR EMBL; S77875; AAD14268.1; JOINED.  
 DR EMBL; AB022177; BAA86929.1; -;  
 DR EMBL; AB022178; BAA86928.1; -;  
 DR PIR; S34486; S34486.  
 DR Genew; HGNC:1440; CALCR.  
 DR MIM; 114131; -;  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PRO2793; HRM\_1.  
 DR PRINTS; PRO0249; GPCRSECRETIN.  
 DR SMART; SM00008; Hoimr; 1.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; 1.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; 1.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECP\_F2\_4; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; signal;  
 KW Alternative splicing; Polymorphism.

FT	SIGNAL	1	22	POTENTIAL.
FT	CHAIN	23	490	CALCITONIN RECEPTOR.
FT	DOMAIN	23	153	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	154	173	1 (POTENTIAL).
FT	DOMAIN	174	196	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	197	216	2 (POTENTIAL).
FT	DOMAIN	217	236	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	237	259	3 (POTENTIAL).
FT	DOMAIN	260	276	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	297	312	4 (POTENTIAL).
FT	DOMAIN	313	336	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	337	359	5 (POTENTIAL).
FT	DOMAIN	360	377	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	378	389	6 (POTENTIAL).
FT	DOMAIN	390	411	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	412	449	7 (POTENTIAL).
FT	DOMAIN	444	449	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	28	73	POLY-ALA.
FT	CARBOHYD	73	73	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	130	130	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	1	47	MISSING (IN ISOFORM 3).
FT	VARSPPLIC	175	190	MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT	VARIANT	463	463	P -> L.
FT	CONFLICT	363	363	/FTID=VAR 003580.
SO	SEQUENCE	490 AA;	57303 MW;	I -> T (IN REF. 3; AAC50301).
				38ECTAA9P07FFSD CRC64;

Query Match 45.1%; Score 37; DB 1; Length 490;  
 Best Local Similarity 40.0%;  
 Matches 6; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Search completed: April 20, 2003, 13:07:40  
 Job time : 4.92105 secs

QY 1 NNTWQFAUQTFT 15  
 DB 124 NNTWQFAUQTFT 138



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds  
(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524d-100

Perfect score: 82

Sequence: 1 NNRIWLPFAKLTGFT 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:  
2: SP bacteria:  
3: SP fungi:  
4: SP human:  
5: SP invertebrate:  
6: SP mammal:  
7: SP mhc:  
8: SP organelle:  
9: SP phage:  
10: SP plant:  
11: SP rodent:  
12: SP virus:  
13: SP vertebrate:  
14: SP unclassified:  
15: SP virus:  
16: SP bacteriophage:  
17: SP archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	63.4	507	10	Q9FY19
2	47	57.3	457	10	Q9ZUB7
3	47	57.3	1161	10	Q9LOD1
4	44	53.7	281	17	Q970D6
5	44	52.4	468	10	Q9CAL5
6	42	51.2	266	16	Q8YUJ9
7	42	51.2	385	3	Q9C1G9
8	42	51.2	439	2	Q93061
9	42	51.2	474	10	Q8W4S2
10	42	51.2	475	10	Q94EH4
11	42	51.2	475	10	Q949Z1
12	42	51.2	492	10	Q9ING3
13	42	51.2	538	12	Q9EX25
14	42	51.2	540	10	Q92699
15	42	51.2	853	3	Q12440
16	41	50.0	220	5	Q9U2G1

17	41	50.0	235	16	Q8XRT9	Q8XRT9 escherichia
18	41	50.0	452	16	Q34521	Q34521 bacillus su
19	40	48.8	75	2	Q9FCW9	Q9FCW9 escherichia
20	40	48.8	246	17	Q9Y9C6	Q9Y9C6 aeropyrum p
21	40	48.8	297	5	Q9W106	Q9W106 drosophila
22	40	48.8	309	5	Q9VAR2	Q9VAR2 drosophila
23	40	48.8	367	10	Q9SR79	Q9SR79 arabidopsis
24	40	48.8	382	10	Q8SP99	Q8SP99 oryza sativ
25	40	48.8	419	10	Q9US84	Q9US84 arabidopsis
26	40	48.8	419	10	Q93Z21	Q93Z21 arabidopsis
27	40	48.8	756	16	P72877	P72877 synechocyst
28	40	48.8	1049	16	Q8Y3H0	Q8Y3H0 ralsronia s
29	40	48.8	1501	3	Q9C2Y4	Q9C2Y4 magnaporthe
30	40	48.8	1501	3	Q9C2Y4	Q9C2Y4 magnaporthe
31	39	47.6	138	10	Q9FYH4	Q9FYH4 arabidopsis
32	39	47.6	217	5	Q95UV7	Q95UV7 biophalari
33	39	47.6	217	16	Q8YFT0	Q8YFT0 brucea me
34	39	47.6	225	3	Q14262	Q14262 schizosach
35	39	47.6	376	2	Q9RG16	Q9RG16 bacteroides
36	39	47.6	399	2	Q9FA51	Q9FA51 rhododactyl
37	39	47.6	485	16	Q9KVV3	Q9KVV3 vibrio chol
38	39	47.6	677	5	Q9XTV1	Q9XTV1 caenorhabdi
39	39	47.6	711	10	Q9SC22	Q9SC22 eremoseyne p
40	39	47.6	1037	16	Q8ZNV7	Q8ZNV7 salmoneilla
41	39	47.6	1037	16	Q8Z4S4	Q8Z4S4 salmoneilla
42	39	47.6	1037	16	Q8XEH2	Q8XEH2 escherichia
43	39	47.6	1277	12	Q9UGJ7	Q9UGJ7 sorghum chl
44	38.5	47.0	652	5	Q8SSV3	Q8SSV3 dictyosteli
45	38	46.3	145	8	Q20657	Q20657 haplochromi

## ALIGNMENTS

RESULT 1	Q9FY19	PRELIMINARY;	PRT;	507 AA.
ID	Q9FY19			
AC	Q9FY19			
DT	01-MAR-2001 (TREMREL. 16, Created)			
DT	01-MAR-2001 (TREMREL. 16, Last sequence update)			
DT	01-JUN-2002 (TREMREL. 21, Last annotation update)			
DE	Pollen major allergen 2 protein precursor.			
GN	JNA2.			
OS	Juniperus ashei (Ozark white cedar).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.			
OX	NCBI_TaxId=13101;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=MALE POLLEN;			
RX	MEDLINE=20403896; PubMed=10944464;			
RA	Yokoyama M., Miyahara M., Shimizu K., Kino K., Tsunoo H.;			
RT	"Purification, identification and cDNA cloning of Jun a 2, the second			
RT	major allergen of mountain cedar pollen."			
RL	Biochem. Biophys. Res. Commun. 275:195-202(2000).			
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES			
CC	(POLYGALACTURONASES).			
CC	EMBL: AJ04653; CAC05582.1; -			
DR	HSSP; P26509; 1BHR.			
DR	InterPro; IPR000743; GH28.			
DR	InterPro; IPR000408; Reg_chr_condens.			
DR	Pfam; PF00295; Glyco_hydro_28; 1.			
DR	PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN_1.			
DR	PROSITE; PS00626; RCCL_2; UNKNOWN_1.			
KW	Cell wall; Glycosidase; Hydrolase; Signal.			
FT	SIGNAL 1			
FT	SEQUENCE 507 AA; 55730 MW; 282E0A5E5958FE5A CRC64;			
QY	1 NNRIWLPFAKLTGFT 14			
Query Match	63.4%;	Score 52;	DB 10;	Length 507;
Best Local Similarity	57.1%;	Pred. No. 0.75;		
Matches	8;	Conservative	4;	Mismatches 2;
		Indels	0;	Gaps 0;

Db 136 NSKIMHFAALITDF 149

## RESULT 2

09ZUE7 PRELIMINARY; PRT; 457 AA.

AC 09ZUE7; 01-MAY-1999 (TREMBLREL. 10, Created)  
 DT 01-MAY-1999 (TREMBLREL. 10, last sequence update)  
 DE 01-MAR-2002 (TREMBLREL. 20, last annotation update)  
 DE F508.2 protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 [1]

SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Liu S., Lenz C.,  
 Li J., Kremenetskaia I., Lueros J., Altafi H., Gonzalez A., Araujo R.,  
 Buehler E., Conn L., Conway A.B., Dunn P., Hansen N., Huizar L.,  
 Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,  
 Ecker J.R., Federpiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F508 sequence."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 [2]

SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 CC EMBL, AC005990; AAC98004.1; -;  
 DR InterPro: IPR000886; ER target.  
 DR InterPro: IPR000743; GH28.  
 DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN 1.  
 DR PROSITE: PS00502; POLYGALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SO SEQUENCE 457 AA; 49269 MW; 0EAA6551E56168 CRC64;

Query Match 57.3%; Score 47; DB 10; Length 457;  
 Best Local Similarity 72.7%; Pred. No. 5.1;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 3 RIMLOFAKLTG 13  
 Db 151 RIMLEFSKMG 161

## RESULT 3

09LOD1 PRELIMINARY; PRT; 1161 AA.

AC 09LOD1; 01-OCT-2000 (TREMBLREL. 15, Created)  
 DT 01-OCT-2000 (TREMBLREL. 15, last sequence update)  
 DE 01-JUN-2002 (TREMBLREL. 21, last annotation update)  
 DE F28C11.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 [1]

SEQUENCE FROM N.A.  
 RP Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
 Kim C., Altafi H., Bei B., Chin C., Chio J., Choi E., Conn L.,  
 Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,  
 Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,

RA Toriumi M., Vaysberg M., Yu G., Federpiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F28C11 from chromosome  
 1.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 [3]

SEQUENCE FROM N.A.  
 RP Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
 Shinn P., Altafi H., Bei B., Chin C., Chio J., Choi E., Conn L.,  
 Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,  
 Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 Toriumi M., Vaysberg M., Yu G., Davis R., Federpiel N., Theologis A.,  
 Ecker J.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 [4]

SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 SEQUENCE FROM N.A.  
 RP Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 Khan S., Kim C., Altafi H., Bei B., Chin C., Chio J., Choi E.,  
 Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N.,  
 Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federpiel N.,  
 Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 CC EMBL, AC007945; AAT79584.1; -;  
 DR InterPro: IPR000886; ER target.  
 DR InterPro: IPR000743; GH28.  
 DR InterPro: IPR002885; PPR.  
 DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
 DR Pfam: PF01535; PPR; 11.  
 DR TIGRfams: TIGR00756; PPR; 9.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN 1.  
 DR PROSITE: PS00502; POLYGALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SO SEQUENCE 1161 AA; 128197 MW; C67DCFBBCBACAB99 CRC64;

Query Match 57.3%; Score 47; DB 10; Length 1161;  
 Best Local Similarity 72.7%; Pred. No. 13;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 3 RIMLOFAKLTG 13  
 Db 824 RIMLEFSKMG 834

## RESULT 4

0970D6 PRELIMINARY; PRT; 281 AA.

AC 0970D6; 01-DEC-2001 (TREMBLREL. 19, Created)  
 DT 01-DEC-2001 (TREMBLREL. 19, last sequence update)  
 DE 01-DEC-2001 (TREMBLREL. 19, last annotation update)  
 DE Hypothetical protein ST1656.  
 GN ST1656.  
 OS Sulfolobus tokodaii.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OC NCBI\_TaxID=111955;  
 [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=JCM 10545 / 7;  
 RC PubMed=11572479;



RA Kawarayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosooyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kuehida N., Oguchi A.,  
 RA Aoki K.-I., Maeda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.,  
 RT "Complete genome sequence of an aerobic thermophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain 7.",  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL; AF000987; BAB6737.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 281 AA; 30125 MW; 465BA5797F383DC CRC64;

Query Match 53.7%; Score 44; DB 17; Length 281;  
 Best Local Similarity 57.1%; Pred. No. 11;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 DB 243 NNTVTFPPPLTGF 256

RESULT 5  
 ID 09CAL5 PRELIMINARY; PRT; 468 AA.  
 AC 09CAL5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative polygalacturonase.  
 GN F24J13.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxId=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,  
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizer L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin B.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luyos J.S., Malt R., Marziani A.,  
 RA Miltischer J., Miranda M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana."  
 RL Nature 408:816-820(2000).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 DR EMBL; AC010796; AAG52465.1; -  
 DR InterPro; IPR000886; ER\_target.  
 DR InterPro; IPR000743; GH28.  
 DR Pfam; PF00285; Glyco\_hydro\_28.1.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 468 AA; 50900 MW; B92D6E5168D7455 CRC64;

Query Match 52.4%; Score 43; DB 10; Length 468;  
 Best Local Similarity 72.7%; Pred. No. 27;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 RNLQFAKLTG 13  
 DB 147 RNLQFSLQ 157

RESULT 6  
 ID 08YQJ9 PRELIMINARY; PRT; 266 AA.  
 AC 08YQJ9;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein A1r3824.  
 GN A1R3824.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OC NCBI\_TaxId=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,  
 RA Matsumoto M., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimizu S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120."  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003594; BAB75523.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 266 AA; 30887 MW; 741814848A5B8C0 CRC64;

Query Match 51.2%; Score 42; DB 16; Length 266;  
 Best Local Similarity 58.3%; Pred. No. 22;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 DB 206 RNLQFSLQ 217

RESULT 7  
 ID 09C1Q9 PRELIMINARY; PRT; 385 AA.  
 AC 09C1Q9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Pteromone receptor CPRalp.  
 GN CPRalp.  
 OS Cryptococcus neoformans (Filobasidiella neoformans).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OC NCBI\_TaxId=5207;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chang Y.C., Kwon-Chung K.J.;  
 RT "Pteromone receptor of Cryptococcus neoformans mating type a."  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF250141; AAK31936.1; -  
 DR InterPro; IPR001499; STE3\_GPCR.  
 DR Pfam; PF02076; STE3\_1.  
 DR PRINTS; PR00899; GPCRSTE3.  
 KW Receptor.  
 SQ SEQUENCE 385 AA; 43122 MW; 146E23DBCF9259E8 CRC64;

Query Match 51.2%; Score 42; DB 3; Length 385;  
 Best Local Similarity 63.6%; Pred. No. 33;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 IWLQFAKLTG 14

Db 37 VWLTFANLQGF 47

## RESULT 8

ID Q93061 PRELIMINARY; PRT; 439 AA.  
 AC Q93061;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Integral membrane protein.  
 GN YIAA2.  
 OS Klebsiella oxytoca.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Klebsiella.  
 CX NCBI\_TaxID=571;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M5A1;  
 RA Darois V.A., Menge K.L., Vande Weghe J.;  
 RT "The ascorbic acid degradation pathway of Klebsiella oxytoca M5a1."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF282849; AKK69524.1;  
 DR InterPro; IPR003662; sub\_transporter.  
 DR Pfam; PF00083; sugar tr; 1.  
 SQ SEQUENCE 439 AA; 47463 MW; AB37CDBB64967257 CRC64;

Query Match 51.2%; Score 42; DB 2; Length 439;  
 Best Local Similarity 58.3%; Pred. No. 37;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Oy 2 NRWLOFAKLTG 13  
 Db 330 NQWLSYALVG 341

## RESULT 9

ID Q8W4S2 PRELIMINARY; PRT; 474 AA.  
 AC Q8W4S2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE At1960590/F8A5.12.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY060568; AAL31197.1;  
 DR InterPro; IPR000743; GH28.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN\_1.  
 SQ SEQUENCE 474 AA; 51964 MW; 9631844B83E209C6 CRC64;

Query Match 51.2%; Score 42; DB 10; Length 474;  
 Best Local Similarity 54.5%; Pred. No. 40;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 5 WLOFAKLTGFT 15  
 Db 159 NKNQWLVFRLDGT 173

Db 139 WIEFTKLGIT 149

## RESULT 10

ID Q94EH4 PRELIMINARY; PRT; 475 AA.  
 AC Q94EH4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE At1948100/F21D18.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 DR EMBL; AF410319; AAK5305.1;  
 DR InterPro; IPR000743; GH28.  
 DR InterPro; IPR004408; Reg chr condens.  
 DR InterPro; IPR001412; RNase-synt\_1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN\_1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN\_1.  
 DR PROSITE; PS00626; RCL1\_2; UNKNOWN\_1.  
 KW Cell wall; Glycosidase; Hydrolase-  
 SQ SEQUENCE 475 AA; 51380 MW; 4A42C4355827D30A CRC64;

Query Match 51.2%; Score 42; DB 10; Length 475;  
 Best Local Similarity 53.3%; Pred. No. 41;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Oy 1 NNRWLOFAKLTGFT 15  
 Db 159 NKNQWLVFRLDGT 173

## RESULT 11

ID Q949Z1 PRELIMINARY; PRT; 475 AA.  
 AC Q949Z1;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative polygalacturonase Pgl.  
 GN F21D18.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,  
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bower L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,  
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Tracy S.E.,

RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full Length cDNA of gene F21D18.18 (GI:8778527).";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 DR EMBL: AY050798; AAK92733.1; -  
 DR InterPro: IPR000743; GH28.  
 DR InterPro: IPR000408; Reg\_chr\_condens.  
 DR InterPro: IPR001412; RNA-synt.1.  
 DR Pfam: PF00295; Glyco\_hydro.28; 1.  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN\_1.  
 DR PROSITE: PS00502; POLYGALACTURONASE; UNKNOWN\_1.  
 DR PROSITE: PS00526; RCC1\_2; UNKNOWN\_1.  
 KM Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 475 AA; 51352 MW; ACB066CF725F2C8F CRC64;  
 Query Match 51.2%; Score 42; DB 10; Length 475;  
 Best Local Similarity 53.3%; Pred. No. 41;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 NNRIMLOFAKLTGFT 15  
 Db 159 NKNQWLVFRLDGT 173  
 RESULT 12  
 ID 09LNG3 PRELIMINARY; PRT; 492 AA.  
 AC 09LNG3;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE F21D18.18;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
 RA Shih P., Altafi H., Bei Q., Chin C., Chiu J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Toriumi M., Vayberg M., Yu G., Federpiet N.A., Theologis A.,  
 RA Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F21D18 from chromosome  
 I.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Chao Q., Shih P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chiu J., Choi E.,  
 RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Theologis A., Toriumi M., Vayberg M., Yu G., Davis R., Federpiet N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 DR EMBL: AC023673; AAF79535.1; -  
 DR InterPro: IPR000743; GH28.  
 DR InterPro: IPR000408; Reg\_chr\_condens.

DR InterPro: IPR001412; RNA-synt.1.  
 DR Pfam: PF00295; Glyco\_hydro.28; 1.  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN\_1.  
 DR PROSITE: PS00502; POLYGALACTURONASE; 1.  
 DR PROSITE: PS00526; RCC1\_2; UNKNOWN\_1.  
 KM Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 492 AA; 53395 MW; 42BAF98B9E30AC1A CRC64;  
 Query Match 51.2%; Score 42; DB 10; Length 492;  
 Best Local Similarity 53.3%; Pred. No. 42;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 NNRIMLOFAKLTGFT 15  
 Db 159 NKNQWLVFRLDGT 173  
 RESULT 13  
 ID 09PX25 PRELIMINARY; PRT; 538 AA.  
 AC 09PX25;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE Nonstructural protein.  
 GN ORF-ALPHA.  
 OS Periplaneta fuliginosa densovirus.  
 OC Viruses; ssDNA viruses; Parvoviridae; Densovirinae; Densovirus.  
 OC NCBI\_TaxID=97344;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamagishi J., Hu Y., Zhang J., Bando H.;  
 RT "Genome organization and mRNA structure of Periplaneta fuliginosa  
 RT densovirus imply alternative splicing involvement in viral gene  
 RT expression.";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Guo H.T., Zhang J.M., Hu Y.Y.;  
 RT "Complete sequence of periplaneta fuliginosa parvovirus.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB028936; BAA82961.1; -  
 DR EMBL: AF192260; AAF04297.1; -  
 KM Nonstructural protein.  
 SQ SEQUENCE 538 AA; 61746 MW; 6E168F8A9236DC5 CRC64;  
 Query Match 51.2%; Score 42; DB 12; Length 538;  
 Best Local Similarity 61.5%; Pred. No. 46;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 NNRIMLOFAKLTG 13  
 Db 181 SNQWLVFRLDGT 193  
 RESULT 14  
 ID 022699 PRELIMINARY; PRT; 540 AA.  
 AC 022699;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative polygalacturonase.  
 GN F8A5.12.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Federpiet N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
 RA Federpiet N.A., Buehler E., Dewar K., Feng J., Kim C.,  
 RA Au M., Araujo R., Brendel V., Buehler E.,

RA Li Y., Oji O., Osborne B.I., Shin P., Sun H., Toriumi M.,  
 RA Vysotskaya V.S., Yu G., Ecker J., Theologos A., Davis R.W.;  
 RT "Genomic sequence of Arabidopsis";  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 CC EMBL; AC002292; AAB71972.1; -.  
 DR EMBL; AC002292; AAB71972.1; -.  
 DR InterPro; IPR000743; GH28.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase  
 SQ SEQUENCE 540 AA; 59631 MW; 532B79A84EBB904 CRC64;  
 QY 5 WLOFAKLGTGFT 15  
 Db 205 WIEFTKLSGIT 215  
 Query Match 51.2%; Score 42; DB 10; Length 540;  
 Best Local Similarity 54.5%; Pred. No. 46;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 15  
 ID Q12440 PRELIMINARY; PRT; 853 AA.  
 AC Q12440;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE ORF YLR127C.  
 GN APC2 OR L3108 OR L3105 OR L9233.13 OR YLR127C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCBI\_TaxID=4932;  
 RX [1]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=FY23/RD005;  
 RA Verhasselt P., Volckaert G.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Delius H.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN RP SEQUENCE FROM N.A.  
 RA Delius H., Hebling U.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RN RP SEQUENCE FROM N.A.  
 RA MIPs;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Pauley A.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
 RA Teich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
 RA Wilson R., Waterston R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 [7]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Waterston R.;

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X89514; CAA61705.1; -.  
 DR EMBL; X91258; CAA62638.1; -.  
 DR EMBL; Z73299; CAA97696.1; -.  
 DR EMBL; U53877; AAB82373.1; -.  
 DR EMBL; Z73300; CAA97698.1; -.  
 DR SGD; S0004117; APC2.  
 DR InterPro; IPR001373; Cullin.  
 DR SMART; SM00182; CULLIN; 1.  
 DR PROSITE; PS50069; CULLIN\_2; 1.  
 SQ SEQUENCE 853 AA; 99977 MW; E559A0E28C784DE2 CRC64;  
 QY 2 NRIWQFAKLGTG 13  
 Db 248 NQYWSQFSKLVG 259  
 Query Match 51.2%; Score 42; DB 3; Length 853;  
 Best Local Similarity 58.3%; Pred. No. 74;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Search completed: April 20, 2003, 13:13:12  
 Job time: 14.6711 secs

GenCore version 5.1.4 ps\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524D-101

Perfect score: 76

Sequence: 1 LQFAKLGTFLMGKG 15

Scoring table: BLOSUM62  
Gapop: 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 73.\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*  
5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	100.0	514	S48730	Cry j II protein -
2	76	100.0	514	JC2498	second major aller
3	54	71.1	514	JC7100	polygalacturonase
4	48	63.2	507	JC7366	Jun a 2 protein -
5	45	59.2	540	B96631	probable polygalac
6	44	57.9	492	C96821	protein F21D18.18
7	43	56.6	194	T35332	hypothetical prote
8	43	56.6	502	H75290	hypothetical prote
9	43	56.6	503	T29870	hypothetical prote
10	43	56.6	519	A96542	probable cytochrom
11	43	56.6	1161	B86368	protein F28C11.9
12	41	53.9	468	H96728	probable polygalac
13	40	52.6	217	AB3416	hypothetical membr
14	40	52.6	231	S50517	hypothetical prote
15	40	52.6	285	I65760	alcohol sulfotrans
16	40	52.6	353	F72394	glutamate 5-kinase
17	40	52.6	358	C72626	hypothetical prote
18	40	52.6	371	F69429	asparagine synthet
19	40	52.6	435	T48628	polygalacturonase
20	40	52.6	567	F70702	ABC transporter si
21	39	51.3	217	AB3432	thiol-disulfide in
22	39	51.3	359	B40354	arylalkylphospha
23	39	51.3	382	D84856	hypothetical prote
24	39	51.3	542	F81075	protein F20B24.8
25	39	51.3	572	F81075	arginyl-tRNA synth
26	39	51.3	572	F81075	arginyl-tRNA synth
27	39	51.3	591	B83457	glyoxylate carboxi
28	39	51.3	1020	B82604	cation efflux syst
29	38	50.0	230	D70718	hypothetical prote

30	38	50.0	312	2	T05348	hypothetical prote
31	38	50.0	329	2	B90879	fermentative D-lac
32	38	50.0	329	2	H85739	fermentative D-lac
33	38	50.0	329	2	G64888	D-lactate dehydrog
34	38	50.0	369	2	S68692	deoxyhypusine synt
35	38	50.0	443	2	T46127	hypothetical prote
36	38	50.0	461	2	S74414	hypothetical prote
37	38	50.0	469	2	G90422	Na+/H+-exchanging
38	38	50.0	594	2	T34855	sugar transport pr
39	38	50.0	639	2	AB3456	probable glyoxylat
40	38	50.0	2793	2	B90784	long-chain-fatty-a
41	38	50.0	2806	2	D85644	hypothetical prote
42	37.5	49.3	244	2	F69830	hypothetical prote
43	37.5	49.3	466	2	T02324	conserved hypotne
44	37.5	49.3	1679	2	S49802	hypothetical prote
45	37	48.7	224	2	T11207	probable membrane
						spap protein - Sal

#### ALIGNMENTS

##### RESULT 1

S48730 Cry j II protein - Japanese cedar

C:Species: Cryptomeria japonica (Japanese cedar)

C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999

C:Accession: S48730

R:Name, M.: Kurose, M.; Torijoe, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.;

FEBS Lett. 353, 124-128, 1994

A:Title: Molecular cloning of the second major allergen, Cry j II, from Japanese cedar

A:Reference number: S48730; MUID:95010777; PMID:7926035

A:Accession: S48730

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-514 <NAM>

A:Cross-references: GB:D37765; NID:9577695; PIDN:BAA07021.1; PID:di007598; PID:9577695

Query Match Best Local Similarity 100.0%; Score 76; DB 2; Length 514;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LQFAKLGTFLMGKG 15

Db 140 LQFAKLGTFLMGKG 154

##### RESULT 2

JC2498 second major allergen Cry j II precursor - Japanese cedar

C:Species: Cryptomeria japonica (Japanese cedar)

C:Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000

C:Accession: JC2498; PC2346; A60147

R:Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994

A:Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese

A:Reference number: JC2498; MUID:94271186; PMID:8002972

A:Accession: JC2498

A:Molecule type: mRNA

A:Residues: 1-514 <KOM>

A:Cross-references: DDBJ:D29772; NID:9506857; PIDN:BAA06172.1; PID:9506858

A:Accession: PC2346

A:Molecule type: protein

A:Residues: 52-61 <KO2>

R:Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matuhasi, T.

Allergy 45, 309-312, 1990

A:Title: Identification of the second major allergen of Japanese cedar pollen.

A:Reference number: A60147; MUID:90342988; PMID:2382797

A:Accession: A60147

A:Molecule type: protein

A:Residues: 55-64 <SAK>

C:Keywords: glycoprotein; pollen

F:1-54/Domains: signal sequence #status predicted <STG>

F:55-460/Product: second major allergen Cry j #status predicted <MAT>

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Query Match

100.0%; Score 76; DB 2; Length 514;

Best Local Similarity 100.0%; Pred. No. 7,8e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 |||||  
 Db 140 LQFAKLTGFTLMGKG 154

## RESULT 3

polygalacturonase Cha o 2 - Japanese cypress

C/Species: Chamaecyparis obtusa (Japanese cypress)

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

C/Accession: JG7100; PC7026

R/Mori, T.; Yokoyama, M.; Komiyama, N.; Okano, M.; Kino, K.

Biochem. Biophys. Res. Commun. 263, 166-171, 1999

A/Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A/Reference number: JG7100; PMID:99417540; PMID:10486272

A/Accession: JG7100

A/Molecule type: mRNA

A/Residues: 1-514 <MOR>

A/Accession: PC7026

A/Molecule type: protein

A/Residues: 51-62 <MO2>

Query Match

71.1%; Score 54; DB 2; Length 514;

Best Local Similarity 73.3%; Pred. No. 0.081;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 |||||  
 Db 140 LQFAKLTGFTLMGKG 154

## RESULT 4

Jun a 2 protein - mountain cedar

C/Species: Juniperus ashei (mountain cedar)

C/Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C/Accession: JG7366; PC7093

R/Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.

Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A/Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A/Reference number: JG7366

A/Accession: JG7366

A/Molecule type: mRNA

A/Residues: 1-507 <YOK>

A/Cross-references: GB:AJ04653

A/Accession: PC7093

A/Molecule type: protein

A/Residues: 55-63 <YOK>

C/Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

C/Keywords: glycoprotein; pollen

Query Match

63.2%; Score 48; DB 2; Length 507;

Best Local Similarity 60.0%; Pred. No. 0.99;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 |||||  
 Db 141 MHFARLTDFTLMGKG 155

## RESULT 5

probable polygalacturonase F8A5.12 (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: B96631

Query Match

57.9%; Score 44; DB 2; Length 492;

Best Local Similarity 60.0%; Pred. No. 5.2;

Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 |||||  
 Db 164 LVFRLDGFPSGKG 178

## RESULT 6

proteins F21D18.16 (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: C96521

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, T.

ansen, N.F.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shum, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; PMID:21016719; PMID:11130712

A/Accession: C96521

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-540 <STO>

A/Cross-references: GB:AE005173; NID:92462753; PIDN:AAW71972.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match

59.2%; Score 45; DB 2; Length 540;

Best Local Similarity 46.7%; Pred. No. 3.7;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 |||||  
 Db 206 IEFKLSGITIGGNG 220

## RESULT 7

hypothetical protein SC5H1.16 - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C/Accession: T35332

R/Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1999

A/Reference number: Z21575

Query Match

57.9%; Score 44; DB 2; Length 492;

Best Local Similarity 60.0%; Pred. No. 5.2;

Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 |||||  
 Db 164 LVFRLDGFPSGKG 178

A:Accession: T35332  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-194 <OLI>  
 A:Cross-references: EMBL:AL049863; PIDN:CA642941.1; GSPDB:GN00070, SCOEDB:SCSH1.16  
 A:Experimental source: strain A3(2)  
 A:Genetics:  
 A:Gene: SCOEDB:SCSH1.16

Query Match 56.6%; Score 43; DB 2; Length 194;  
 Best Local Similarity 63.6%; Pred. No. 2.9;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 KUTGFTLMGK 15  
 DB 37 RITGFTLMGK 47

RESULT 8  
 H75290  
 hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: H75290  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 S.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: H75250  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-502 <WHI>  
 A:Cross-references: GB:AE002061; GB:AE000513; NID:g6460095; PIDN:AAF11844.1; PID:g646010  
 A:Experimental source: strain R1  
 A:Genetics:  
 A:Gene: DR2286  
 A:Map position: 1

Query Match 56.6%; Score 43; DB 2; Length 502;  
 Best Local Similarity 47.1%; Pred. No. 8;  
 Matches 8; Conservative 7; Mismatches 0; Indels 2; Gaps 1;

OY 1 LQPAKLTGF--TLMGK 15  
 DB 287 VEFARLGGYRLTLGK 303

RESULT 9  
 T29870  
 hypothetical protein F32B5.6 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T29870  
 R:ledwith, J.; Graves, T.; Biewald, T.  
 submitted to the EMBL Data Library, May 1997  
 A:Description: The sequence of C. elegans cosmid F32B5.  
 A:Reference number: Z20702  
 A:Accession: T29870  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1503 <LED>  
 A:Cross-references: EMBL:AF003148; PIDN:AAB54212.1; GSPDB:GN00019; CESP:F32B5.6  
 A:Experimental source: strain Bristol N2; clone F32B5  
 A:Genetics:  
 A:Gene: CESP:F32B5.6  
 A:Map position: 1  
 A:Initons: 33/2; 71/1; 210/3; 249/3; 299/3; 358/2; 398/2; 415/1; 453/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F32B5.6

Query Match 56.6%; Score 43; DB 2; Length 503;  
 Best Local Similarity 53.8%; Pred. No. 8.1;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 QPAKLTGFTLMGK 14  
 DB 429 KFAEFAAGFTVLGK 441

RESULT 10

A96542  
 probable cytochrome P450 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 20-Apr-2001  
 C:Accession: A96542  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: A96542  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-519 <STO>  
 A:Cross-references: GB:AE005173; NID:g11054637; PIDN:AAG27882.1; GSPDB:GN00141  
 A:Genetics:  
 A:Gene: F1706.8  
 A:Map position: 1  
 C:Superfamily: human cytochrome P450 CY2D6; cytochrome P450 homology  
 C:Keywords: heme; iron; metalloprotein  
 F:58/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 56.6%; Score 43; DB 2; Length 519;  
 Best Local Similarity 61.5%; Pred. No. 8.3;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 QPAKLTGFTLMGK 14  
 DB 239 QFSKLWGISLFGK 251

RESULT 11

B86368  
 protein F28C11.9 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B86368  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: B86368  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1161 <STO>  
 A:Cross-references: GB:AE005172; NID:g8778576; PIDN:AAF79584.1; GSPDB:GN00141  
 A:Genetics:  
 A:Gene: F28C11.9  
 A:Map position: 1

Query Match 56.6%; Score 43; DB 2; Length 1161;  
 Best Local Similarity 53.3%; Pred. No. 20;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 LQFALGTGTLMGKG 15  
 ||:|||||  
 Db 827 LQFSLKGVVFGQNG 841

## RESULT 12

H96728

probable polygalacturonase F24J13.7 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: H96728

R:Theologas, A.; Becker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzilli,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H96728

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-468 <STO>

A:Cross-references: GB:AE005173; NID:g6175142; PIDN:AAF04869.1; GSPDB:GN00141

A:Gene: F24J13.7

A:Map position: 1

Query Match 53.9%; Score 41; DB 2; Length 468;

Best Local Similarity 53.3%; Pred. No. 17;

Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQFALGTGTLMGKG 15

||:|||||

Db 150 LQFSLKGVVFGQNG 164

## RESULT 13

AB3416

hypothetical membrane spanning protein BMEI1312 [imported] - Brucella melitensis (strain

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AB3416

R:DeVechio, V.G.; Kaprat, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AB3416

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-217 <KUR>

A:Cross-references: GB:AE008917; PIDN:AA152493.1; PID:g17983302; GSPDB:GN00190

A:Experimental source: strain 16M

A:Gene: BMEI1312

A:Map position: 1

## RESULT 14

Query Match 52.6%; Score 40; DB 2; Length 217;

Best Local Similarity 69.2%; Pred. No. 12;

Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 FAKLTGFTLMGKG 15

||:|||||

Db 125 FALMGFTLMGKG 137

S50517

hypothetical protein YEL072w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 19-Apr-2002

C:Accession: S50517

R:Dietrich, F.S.

A:Description: The sequence of S. cerevisiae cosmid 9669, 8334, 8199, and lambda clone

A:Reference number: S50434

A:Accession: S50517

A:Molecule type: DNA

A:Residues: 1-231 <DIR>

A:Cross-references: EMBL:U18795; NID:g603241; PID:g603246; GSPDB:GN00005; MIPS:YEL072w

A:Gene: MIPS:YEL072w

A:Cross-references: GSD:S0000798

A:Map position: 5L

Query Match 52.6%; Score 40; DB 2; Length 231;

Best Local Similarity 63.6%; Pred. No. 12;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 KLTGFTLMGKG 15

||:|||||

Db 137 EITGFTSMGNG 147

## RESULT 15

I65760

alcohol sulfoltransferase (EC 2.8.2.2) - rat

N:Alternate names: hydroxysteroid sulfoltransferase; senescence marker protein 2B

C:Species: Rattus norvegicus (Norway rat)

C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 18-Jun-1999

C:Accession: I65760; I52407

R:Watabe, T.; Ogura, K.; Satsukawa, M.; Okuda, H.; Hiratsuka, A.

Chem. Biol. Interact. 92, 87-105, 1994

A:Title: Molecular cloning and functions of rat liver hydroxysteroid sulfoltransferase

A:Reference number: I52849; MUID:94306585; PMID:8033273

A:Accession: I65760

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-285 <RBS>

A:Cross-references: GB:D14989; NID:g303801; PIDN:BA03634.1; PID:g303802

R:Song, C.S.; Kim, J.M.; Roy, A.K.; Chatterjee, B.

Biochemistry 29, 542-551, 1990

A:Title: Structure and regulation of the senescence marker 2 gene promoter gene promo:

A:Reference number: I52408; MUID:90148982; PMID:2302387

A:Accession: I52407

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 2-68 <RBS>

A:Cross-references: GB:M29302; NID:g206995; PIDN:AAA42152.1; PID:g206996

C:Gene: I65760

C:Introns: 46/1

C:Superfamily: alcohol sulfoltransferase

C:Keywords: sulfoltransferase

Query Match 52.6%; Score 40; DB 2; Length 285;

Best Local Similarity 80.0%; Pred. No. 16;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 LTGFTLMGKG 15

||:|||||

Search completed: April 20, 2003, 13:15:52

Job time : 7.07895 secs



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-101

Perfect score: 76

Sequence: 1 LQPAKLTGFTLMGK 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	100.0	514	1	MPA2_CRYJA
2	40	52.6	231	1	YE12_YEAST
3	40	52.6	285	1	SPAP_SALT
4	40	52.6	353	1	PROB_THEMA
5	40	52.6	367	1	Y095_SYNY3
6	39	51.3	358	1	PONI_RABIT
7	38	50.0	230	1	Y966_MYCTU
8	38	50.0	329	1	LDHD_ECOLI
9	38	50.0	369	1	DHYS_HUMAN
10	38	50.0	759	1	RRP3_TAKTE
11	37.5	49.3	1679	1	YMF9_YEAST
12	37	48.7	224	1	SPAP_SALT
13	37	48.7	224	1	SPAP_SALT
14	37	48.7	244	1	SPAP_SALT
15	37	48.7	280	1	LEPA_ALCEU
16	37	48.7	280	1	LEPA_ALCEU
17	37	48.7	287	1	Y193_HAETN
18	37	48.7	397	1	Y064_MYCTU
19	37	48.7	407	1	MOSC_RHIME
20	37	48.7	438	1	DOCTA_BRAJA
21	37	48.7	911	1	TEB1_NEIMO
22	37	48.7	915	1	TEB1_NEIMO
23	36	47.4	160	1	NRD1_STEPPY
24	36	47.4	210	1	NRD1_STEPPY
25	36	47.4	230	1	DOCTA_BRAJA
26	36	47.4	315	1	YB52_SYNY3
27	36	47.4	331	1	YB52_SYNY3
28	36	47.4	347	1	ADH2_KLUMA
29	36	47.4	608	1	YLPK_CAREL
30	36	47.4	611	1	GLMS_NETMA
31	36	47.4	611	1	GLMS_NETMA
32	36	47.4	785	1	Y043_MERYA
33	36	47.4	810	1	YP07_YEAST

## ALIGNMENTS

34	36	47.4	1120	1	MYD_RICPR	00555	ricetteia
35	36	47.4	1582	1	YF01_PALSO	08x702	raibsonia s
36	35	46.1	239	1	CYRH_RHIV	P7648	rhizobium 1
37	35	46.1	247	1	YBGI_ECOLI	P75743	escherichia
38	35	46.1	305	1	RNPI_CAEBL	Q10667	caenorhabdi
39	35	46.1	329	1	E2B1_METUA	057896	methanococc
40	35	46.1	375	1	SRF_CHICK	Q90718	gallus gall
41	35	46.1	397	1	PGIR_BRANA	P35337	brassica na
42	35	46.1	462	1	PRYC_ECOLI	P33059	escherichia
43	35	46.1	506	1	AMID_STNY3	Q55424	synchocyst
44	35	46.1	513	1	YHGX_BACSU	P54608	bacillus su
45	35	46.1	569	1	STLI_YEAST	P39932	saccharomyc

ALIGNMENTS

RESULT 1

MPA2\_CRYJA

MPA2\_CRYJA

STANDARD;

PRT;

514 AA.

AC

P4312;

01-NOV-1995 (Rel. 32, Created)

01-NOV-1995 (Rel. 40, Last annotation update)

16-OCT-2001 (Rel. 40, Last annotation update)

DE

Possible polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)

DS

(Major pollen allergen Cry j 2) (Cry j II).

OS

Cryptomeria japonica (Japanese cedar).

OC

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Cryptomeria.

OX

Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.

NCB1\_Taxid=3369;

[1]

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RP

TISSUE=Pollen;

RA

MEDLINE=95010777; PubMed=7926035;

RA

Namba M., Kurose M., Torioe K., Hino K., Taniguchi Y., Fukuda S.,

RA

Usui M., Kurimoto M.;

RA

"Molecular cloning of the second major allergen, Cry j II, from

RT

Japanese cedar pollen.";

RL

FEBS Lett. 353:124-128(1994).

[2]

SEQUENCE FROM N.A.

RP

TISSUE=Pollen;

RA

MEDLINE=94271186; PubMed=8002972;

RA

Komiyama N., Some T., Shimizu K., Morikubo K., Kino K.;

RT

"cDNA cloning and expression of Cry j II the second major allergen of

RL

Japanese cedar pollen.";

RL

Biochem. Biophys. Res. Commun. 201:1021-1028(1994).

[3]

SEQUENCE OF 55-64.

RP

MEDLINE=90342988; PubMed=2382797;

RX

Sakaguchi M., Inoue S., Tanai M., Ando S., Usui M., Matubasi T.;

RA

"Identification of the second major allergen of Japanese cedar

RT

pollen.";

RL

Allergy 45:309-312(1990).

CC

-1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-

CC

galactosiduronic linkages in pectate and other galacturonans.

CC

-1- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).

CC

-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES

CC

(POLYGALACTURONASES).

CC

-----

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC

the European Bioinformatics Institute. There are no restrictions on its

CC

use by non-profit institutions as long as its content is in no way

CC

modified and this statement is not removed. Usage by and for commercial

CC

entities requires a license agreement (See <http://www.isb-sib.ch/announce/isb-sib.ch>).

CC

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC

-----

DR

EMBL; D37765; BAA07021.1; -

DR

EMBL; D29772; BAA06172.1; -

DR

HSP; P26509; IBHE.

DR

InterPro; IPR00743; GH28.

DR

Pfam; PF00295; Glyco\_Hyd2\_28; 1.

DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 KW Hydrolyase; Glycosidase; Cell wall; Signal; Zymogen; Fruit ripening;  
 KM Amyloplast; Glycoprotein; Allergen.  
 FT SIGNAL 1 22  
 FT PROPEP 23 45  
 FT CHAIN 46 433  
 FT PROPEP 434 514  
 FT ACT SITE 278 278  
 FT CARBOHYD 460 460  
 FT CARBOHYD 472 472  
 FT CARBOHYD 5 5  
 FT CONFLICT 12 12  
 FT CONFLICT 34 35  
 FT CONFLICT 37 37  
 FT CONFLICT 88 88  
 FT CONFLICT 98 98  
 FT CONFLICT 451 451  
 FT CONFLICT 454 454  
 FT CONFLICT 504 504  
 FT CONFLICT 507 507  
 SQ SEQUENCE 514 AA; 56645 MW; 624611C3FABD6302 CRC64;  
 Query Match 100.0%; Score 76; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LGFTLMGKG 15  
 DB 140 LGFTLMGKG 154

RESULT 2  
 ID YE12 YEAST STANDARD; PRT; 231 AA.  
 AC P39975;  
 DT 01-PEB-1995 (Rel. 31, Created)  
 DT 01-PEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 26.8 kDa protein in DLD3 5' region.  
 GN YEL072W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
 OK NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1288c / AB972;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
 RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,  
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,  
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U18795; AAB65015.1; -  
 DR SGD; S0000798; YEL072W.  
 KM Hypothetical protein.  
 SQ SEQUENCE 231 AA; 26815 MW; 1C6A1A3CE19D5B1C CRC64;  
 Query Match 52.6%; Score 40; DB 1; Length 231;  
 Best Local Similarity 63.6%; Pred. No. 5.3;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 LGFTLMGKG 15  
 DB 137 LGFTLMGKG 147

RESULT 3  
 ID SUH3\_RAT STANDARD; PRT; 285 AA.  
 AC P50235;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid  
 DE sulfotransferase) (ST) (ST-60).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OK NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RA MEDLINE=94306585; PubMed=8033273;  
 RX Matabe T., Ogura K., Satsukawa M., Okuda H., Hiratsuka A.;  
 RT "Molecular cloning and functions of rat liver hydroxysteroid  
 RT sulfotransferases catalysing covalent binding of carcinogenic  
 RT polycyclic arylmethanols to DNA."  
 RL Chem. Biol. Interact. 92:87-105(1994).  
 CC -1- FUNCTION: CATALYZES SULFATION OF HYDROXYSTERIODS AND XENOBIOTICS  
 CC (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =  
 CC adenosine 3',5'-bisphosphate + an alkyl sulfate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- INDUCTION: INDUCED BY ESTROGENS AND SUPPRESSED BY ANDROGENS.  
 CC EXPRESSION IS UNDER THE INFLUENCE OF PITUITARY GROWTH HORMONE AND  
 CC THYROID HORMONE.  
 CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D14989; BAA03634.1; -  
 DR HSSP; P50224; ICM.  
 DR InterPro; IPR000863; Sulfotransferase.  
 DR Pfam; PF00685; Sulfotransferase; 1.  
 DR ProDom; PD001218; Sulfotransferase; 1.  
 KW Transferase; Steroid metabolism.  
 FT BINDING 249 255  
 FT BINDING 249 255  
 SQ SEQUENCE 285 AA; 35531 MW; 5ACAS6A859F9D4D5 CRC64;  
 Query Match 52.6%; Score 40; DB 1; Length 285;  
 Best Local Similarity 80.0%; Pred. No. 6.7;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 LGFTLMGKG 15  
 DB 240 LGFTLMGKG 249

RESULT 4  
 ID PROB\_THEMA STANDARD; PRT; 353 AA.  
 AC Q9WYD0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glutamate 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK).  
 GN PROB OR TW0294.  
 OS Thermotoga maritima.

OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;  
 CC Thermotogaceae; Thermotoga.  
 CC NCBI\_TaxID=2336;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=MSB8 / DSM 3109;  
 CC MEDLINE=9287316; PubMed=10360571;  
 CC Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 CC Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 CC McDonald L., Utterback T.R., Malek J.A., Lither K.D., Garrett M.M.,  
 CC Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 CC Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 CC Salzberg S.L., Smith H.O., Venter J.C., Fraser C.W.;  
 CC "Evidence for lateral gene transfer between Archaea and Bacteria from  
 CC genome sequence of *Thermotoga maritima*.";  
 CC RT Nature 399:323-329 (1999).  
 CC -1- FUNCTION: Catalyzes the transfer of a phosphate group to glutamate  
 CC to form glutamate 5-phosphate which rapidly cyclizes to 5-  
 CC oxoprolinone.  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-  
 CC phosphate.  
 CC -1- PATHWAY: Proline biosynthesis; first step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMATE 5-KINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE001711; AAD35382.1; -  
 CC TIGR; TM0294; -  
 CC InterPro; IPR001048; Aa\_kinase.  
 CC InterPro; IPR001057; Glu\_5kinase.  
 CC InterPro; IPR002478; PUA.  
 CC Pfam; PF00696; aak\_kinase; 1.  
 CC Pfam; PF01472; PUA; 1.  
 CC PRINTS; PR00474; GLU5KINASE.  
 CC SMART; SM00359; PUA; 1.  
 CC TIGRFAMs; TIGR01027; prob; 1.  
 CC PROSITE; PS00902; GLUTAMATE\_5\_KINASE; 1.  
 CC Transferrase; Kinase; Prolinase biosynthesis; Complete proteome.  
 CC KW TRANSFERASE; KINASE; PROLINE BIOSYNTHESIS; COMPLETE PROTEOME.  
 CC SQ SEQUENCE 353 AA; 38320 MW; 72E3171CB9B9760 CRC64;  
 CC -----  
 CC Query Match 52.6%; Score 40; DB 1; Length 353;  
 CC Best Local Similarity 58.3%; Pred. No. 8.3;  
 CC Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 CC QY 4 AKLTGFTLMGK 15  
 CC | : ||| : |||  
 CC Db 49 ARAAGFTYLGKG 60  
 CC -----  
 CC RESULT 5  
 CC Y095\_SYNY3 STANDARD; PRT; 567 AA.  
 CC AC Q55884;  
 CC DT 15-JUL-1998 (Rel. 36, Created)  
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 CC DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Hypothetical protein sli0095.  
 CC GN Sli0095.  
 CC OS *Synechocystis* sp. (strain PCC 6803).  
 CC OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 CC NCBI\_TaxID=1148;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=96127529; PubMed=8590279;  
 CC Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 CC Sugita M., Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 644 to 924 of the genome.";  
 CC DNA Res. 2:153-166 (1995).  
 CC -1- SIMILARITY: BELONGS TO THE ABC1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D64004; BAA10646.1; -  
 CC DR InterPro; IPR004147; ABC1.  
 CC DR Pfam; PF03109; ABC1; 1.  
 CC KW Hypothetical protein; Complete proteome.  
 CC SQ SEQUENCE 567 AA; 64990 MW; 1D52C6A6934CA24 CRC64;  
 CC -----  
 CC Query Match 52.6%; Score 40; DB 1; Length 567;  
 CC Best Local Similarity 75.0%; Pred. No. 14;  
 CC Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 CC QY 1 LQFAKLTGFTLM 12  
 CC | : ||| : |||  
 CC Db 508 LIFACLTGFTLL 519  
 CC -----  
 CC RESULT 6  
 CC P0N1\_RABIT STANDARD; PRT; 358 AA.  
 CC AC P27170;  
 CC DT 01-AUG-1992 (Rel. 23, Created)  
 CC DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 CC DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Serum paroxonase/arylesterase 1 (EC 3.1.1.2) (EC 3.1.1.2) (PON 1)  
 CC (Serum aryl dialkylphosphatase 1) (A-esterase 1) (Aromatic esterase 1).  
 CC GN PON1 OR PON.  
 CC OS *Oryctolagus cuniculus* (Rabbit).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 CC NCBI\_TaxID=9986;  
 CC [1]  
 CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 CC RP TISSUE=Liver;  
 CC RX MEDLINE=92031445; PubMed=1657140;  
 CC RA Haselt C., Richter R.J., Humbert R., Chapline C., Crabb J.W.,  
 CC Omiecinski C.J., Furlong C.E.;  
 CC "Characterization of cDNA clones encoding rabbit and human serum  
 CC paroxonase: the mature protein retains its signal sequence.";  
 CC Biochemistry 30:10141-10149 (1991).  
 CC [2]  
 CC SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 CC RX MEDLINE=93345100; PubMed=8393745;  
 CC RA Furlong C.E., Costa L.G., Haselt C., Richter R.J.,  
 CC Sundstrom J.A., Adler D.A., Distche C.M., Omiecinski C.J.,  
 CC Chapline C., Crabb J.W.;  
 CC "Human and rabbit paroxonases: purification, cloning, sequencing,  
 CC mapping and role of polymorphism in organophosphate detoxification.";  
 CC Chem. Biol. Interact. 87:35-48 (1993).  
 CC [3]  
 CC CHARACTERIZATION, AND SEQUENCE OF 1-20.  
 CC RX MEDLINE=92031444; PubMed=1718413;  
 CC RA Furlong C.E., Richter R.J., Chapline C., Crabb J.W.;  
 CC "Purification of rabbit and human serum paroxonase.";  
 CC Biochemistry 30:10133-10140 (1991).  
 CC -1- FUNCTION: HYDROLYZES THE TOXIC METABOLITES OF A VARIETY OF  
 CC ORGANOPHOSPHORUS INSECTICIDES. CAPABLE OF HYDROLYZING A BROAD  
 CC SPECTRUM OF ORGANOPHOSPHATE SUBSTRATES AND A NUMBER OF AROMATIC  
 CC CARBOXYLIC ACID ESTERS.  
 CC -1- CATALYTIC ACTIVITY: Aryl dialkyl phosphate + H(2)O = dialkyl  
 CC phosphate + an aryl alcohol.



RA Yamamoto Y., Horiuchi T.;  
 RT "A 570-Kb DNA sequence of the Escherichia coli K-12 genome  
 RL corresponding to the 28.0-40.1 min region on the linkage map.";  
 RN DNA Res. 3:363-377(1996).  
 [4]  
 RP CHARACTERIZATION AS AN HEAT-SHOCK GENE.  
 RX MEDLINE=93352431; PubMed=849564;  
 RA Chuang S.B., Blatner F.R.;  
 RT "Characterization of twenty-six new heat shock genes of Escherichia  
 colli.";  
 RL J. Bacteriol. 175:5242-5252(1993).  
 CC -1- FUNCTION: FERMENTATIVE LACTATE DEHYDROGENASE.  
 CC -1- CATALYTIC ACTIVITY: (R)-lactate + NAD(+) = pyruvate + NADH.  
 CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID  
 DEHYDROGENASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U36928; AAB51772.1; -  
 DR EMBL: AB000235; AAC74462.1; -  
 DR EMBL: D90775; BAI14985.1; -  
 DR EMBL: D90776; BAI14990.1; -  
 DR HSSP: P30901; 2DLD.  
 DR ECO2DBASE: D033.4; 6TH EDITION.  
 DR EcGene: EG3186; 1dbA.  
 DR InterPro: IPR002162; D\_2hyd.ac.dh.  
 DR Pfam: PF00389; 2-Hacid.DH; 1.  
 DR Pfam: PF02826; 2-Hacid.DH; 1.  
 DR PROSITE: PS00065; D\_2\_HYDROXYACID\_DH\_1; 1.  
 DR PROSITE: PS00670; D\_2\_HYDROXYACID\_DH\_2; 1.  
 DR PROSITE: PS00671; D\_2\_HYDROXYACID\_DH\_3; 1.  
 KM Oxidoreductase; NAD; Heat shock; Complete proteome.  
 FT NP BIND 146 174 NAD (By SIMILARITY).  
 FT ACT SITE 234 263 SUBSTRATE-BINDING (By SIMILARITY).  
 FT ACT SITE 263 295 BY SIMILARITY.  
 FT ACT SITE 295 295 BY SIMILARITY.  
 SQ SEQUENCE 329 AA; 36534 MW; EF85419986438D6D CRC64;  
 Query Match 50.0%; Score 38; DB 1; Length 329;  
 Best Local Similarity 77.8%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 LTGFTLMGK 14  
 DB 137 LTGFTMGK 145  
 RESULT 9  
 DHYS\_HUMAN STANDARD; PRT; 369 AA.  
 ID P49366; Q13276; Q13184;  
 AC P49366; Q13276; Q13184;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Deoxyhypusine synthase (EC 2.5.1.46) (DHS).  
 GN DHS OR DS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBT\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95403436; PubMed=7673224;  
 RA Joe Y.A., Wolfe E.C., Park M.H.;  
 RT "Cloning and expression of human deoxyhypusine synthase cDNA.  
 RT Structure-function studies with the recombinant enzyme and mutant  
 RT proteins.";

RL J. Biol. Chem. 270:22386-22392(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=96140738; PubMed=8549832;  
 RA Bevec D., Kappel B., Jakse H., Geonga R., Hauber J., Klier H.,  
 RA Steinkasserer A.;  
 RT "Molecular characterization of a cDNA encoding functional human  
 RT deoxyhypusine synthase and chromosomal mapping of the corresponding  
 RT gene locus.";  
 RL FEBS Lett. 378:195-198(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96201991; PubMed=8615810;  
 RA Yan Y.P., Tao Y., Chen K.Y.;  
 RT "Molecular cloning and functional expression of human deoxyhypusine  
 RT synthase cDNA based on expressed sequence tag information.";  
 RL Biochem. J. 315:429-434(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97264341; PubMed=9110174;  
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
 RA Ricchetti J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
 RT "Large-scale concatenation cDNA sequencing.";  
 RL Genome Res. 7:353-358(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98332729; PubMed=9666110;  
 RA Mantuano E., Trettel F., Olsen A.S., Lemm G., Frontali M.,  
 RA Jodice C.;  
 RT "Localization and genomic structure of human deoxyhypusine synthase  
 RT gene on chromosome 19p13.2-distal 19p13.1.";  
 RL Gene 215:153-157(1998).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain, and Lung;  
 RA Strausberg R.;  
 RN Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=98154315; PubMed=9493264;  
 RA Liao D.-I., Wolfe E.C., Park M.H., Davies D.R.;  
 RT "Crystal structure of the NAD complex of human deoxyhypusine  
 RT synthase: an enzyme with a ball-and-chain mechanism for blocking the  
 RT active site.";  
 RL Structure 6:23-32(1998).  
 CC -1- FUNCTION: CATALYZES THE NAD-DEPENDENT OXIDATIVE CLEAVAGE OF  
 CC SPERMIDINE AND THE SUBSEQUENT TRANSFER OF THE BUTYLAMINE MOIETY OF  
 CC OF THE EIF-5A PRECURSOR PROTEIN TO FORM THE INTERMEDIATE  
 CC DEOXYHYPUSINE RESIDUE.  
 CC -1- CATALYTIC ACTIVITY: [eIF5A-precursor]-lysine + spermidine =  
 CC [eIF5A-precursor]-deoxyhypusine + propan-1,3-diamine.  
 CC -1- COFACTOR: NAD.  
 CC -1- PATHWAY: Hypusine biosynthesis; first step.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM  
 CC IS INACTIVE.  
 CC -1- SIMILARITY: BELONGS TO THE DEOXYHYPUSINE SYNTHASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L39068; AAA86282.1; -  
 DR EMBL: U40579; AAA96151.1; -  
 DR EMBL: U32178; AAB02179.1; -  
 DR EMBL: U26266; AAB02175.1; -

DR EMBL; U79262; AAB50208.1; --  
 DR EMBL; AJ001701; CA04940.1; --  
 DR EMBL; AJ001702; CA04940.1; JOINED.  
 DR EMBL; AJ001703; CA04940.1; JOINED.  
 DR EMBL; AJ001704; CA04940.1; JOINED.  
 DR EMBL; BC000333; AAH00333.1; --  
 DR EMBL; BC014016; AAH14016.1; --  
 DR PDB; 1DHS; 25-FEB-98.  
 DR Genew; HGNC:2869; DHS.  
 DR MIM; 600944; --  
 DR InterPro; IPR002773; DS.  
 DR Pfam; PF01916; DS; 1.  
 DR ProDom; PD007730; DS; 1.  
 DR TIGRFAMs; TIGR00321; dhys; 1.  
 KM Hypusine biosynthesis; Transferase; NAD; Alternative splicing;  
 3d-structure.  
 FT VARSPIC 262 308 MISSING (IN SHORT ISOCORR).  
 FT CONFLICT 11 11 A -> R (IN REF. 2 AND 3).  
 FT CONFLICT 13 14 AL -> R (IN REF. 3).  
 FT CONFLICT 85 85 A -> G (IN REF. 3).  
 FT CONFLICT 196 196 T -> I (IN REF. 3).  
 FT CONFLICT 199 199 V -> A (IN REF. 3).  
 FT CONFLICT 220 220 V -> A (IN REF. 3).  
 FT CONFLICT 296 297 MR -> SG (IN REF. 3).  
 FT CONFLICT 311 311 E -> EE (IN REF. 3).  
 SQ SEQUENCE 369 AA; 40970 MW; 5314FED620AC9BE7 CRC64;

Query Match 50.0%; Score 38; DB 1; Length 369;  
 Best Local Similarity 53.8%; Pred. No. 20;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 FAKLTGFTLMGKG 15  
 Db 272 FAKTGFTLMGKG 284

RESULT 10  
 RRP3\_IATKE STANDARD; PRT; 759 AA.  
 ID RRP3\_IATKE  
 AC P18884;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE RNA-directed RNA polymerase subunit p3 (EC 2.7.7.48) (Polymerase basic  
 protein 2) (PB2).  
 OS Influenza A virus (strain A/Kiev/59/79).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza A viruses; Influenzavirus A.  
 CC NCBI\_Taxid=11422;  
 RX MEDLINE=8049805; PubMed=3675640;  
 RA Petrov N.A., Netesov S.V., Golovin S.Y., Mamaeva N.V.,  
 RA Mamaev I.V., Sivododova G.F., Petrenko V.A., Vasilenko S.K.;  
 RT "Primary structure of the full-length DNA copy of the influenza virus  
 A/Kiev/59/79 (H1N1) PB2 gene";  
 RL Bioorg. Khim. 13:915-920(1987).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 (RNA) (N).  
 CC -1- SUBUNIT: INFLUENZA RNA POLYMERASE IS COMPOSED OF THREE SUBUNITS:  
 CC P1 (OR PB1), P2 (OR PA), AND P3 (OR PB2).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA VIRUSES POLYMERASE PB2  
 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M38277; AAA43653.1; --

DR PIR; J00411; P31VAK.  
 DR InterPro; IPR001591; Flu\_PB2.  
 DR Pfam; PF00604; Flu\_PB2; 1.  
 DR ProDom; PD001667; Flu\_PB2; 1.  
 KM Transferase; RNA-directed RNA polymerase.  
 SQ SEQUENCE 759 AA; 86000 MW; 9272580CEA705AFC CRC64;

Query Match 50.0%; Score 38; DB 1; Length 759;  
 Best Local Similarity 46.7%; Pred. No. 43;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LOPAKLTGFTLMGKG 15  
 Db 686 VESAVLRGFTLMGKG 700

RESULT 11  
 YMF9\_YEAST STANDARD; PRT; 1679 AA.  
 ID YMF9\_YEAST  
 AC 004958;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 187.1 kDa protein in OG31-CNA2 intergenic region.  
 GN YML059C OR YMF958.03C.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CC NCBI\_Taxid=4932;  
 RX STRAIN=5288C / AB972;  
 RA Devlin K., Churcher C., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE YMF0028 (SMS) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; Z46729; CA86716.1; --  
 DR SGD; S0004524; YML059C.  
 DR InterPro; IPR002641; Patatin.  
 DR InterPro; IPR001423; YPF0028.  
 DR InterPro; IPR000595; CNMP\_binding.  
 DR Pfam; PF00027; CNMP\_binding; 2.  
 DR Pfam; PF01734; Patatin; 1.  
 DR SMART; SM00100; CNMP; 1.  
 DR PROSITE; PSS0042; CNMP\_BINDING\_3; 2.  
 DR PROSITE; PSS01237; YPF0028; 1.  
 DR Hypothetical protein; Transmembrane.  
 FT TRANSMEM 50 70 POTENTIAL.  
 FT TRANSMEM 104 124 POTENTIAL.  
 SQ SEQUENCE 1679 AA; 187132 MW; 550FFCD4ACAF8B25 CRC64;

Query Match 49.3%; Score 37.5; DB 1; Length 1679;  
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
 Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 1 LOPAKLTGFTLMGKG 15  
 Db 1280 IQFS-LITFTLMGKG 1293

RESULT 12  
 SPAP\_SALT1 STANDARD; PRT; 224 AA.  
 ID SPAP\_SALT1  
 AC Q56023;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Surface presentation of antigens protein spap.  
 GN SPAP OR STY3013.  
 OS Salmonella typhi.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 NCBI\_TaxID=601;  
 RX MEDLINE=95365346; PubMed=7638176;  
 RA Li J., Ochman H., Groisman E.A., Boyd E.F., Solomon F., Nelson K.,  
 RA Selander R.K.,  
 RT "Relationship between evolutionary rate and cellular location among  
 RT the Inv/Spa invasion proteins of Salmonella enterica."  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7252-7256(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RC MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18."  
 RL Nature 413:848-852(2001).  
 CC -1- FUNCTION: INVOLVED IN A SECRETORY PATHWAY RESPONSIBLE FOR THE  
 CC SURFACE PRESENTATION OF DETERMINANTS NEEDED FOR THE ENTRY OF  
 CC SALMONELLA SPECIES INTO MAMMALIAN CELLS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE FLIP/MOPC/SPAP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U29363; AAC43861.1; -  
 DR EMBL: AL627276; CAD05997.1; -  
 DR InterPro: IPR002039; FLIP.  
 DR Pfam: PF00813; FLIP; 1.  
 DR PRINTS: PR01302; TYPE3IMPROT.  
 DR PRODOM: PD002586; FLIP; 1.  
 DR TIGRFAm: TIGR01102; yscr, 1.  
 DR PROSITE: PS01060; FLIP\_1; 1.  
 DR PROSITE: PS01061; FLIP\_2; 1.  
 KW Virulence; Transmembrane; Complete proteome.  
 FT TRANSMEM 8  
 FT TRANSMEM 55  
 FT TRANSMEM 161  
 FT TRANSMEM 185  
 FT TRANSMEM 205  
 FT TRANSMEM 224  
 FT TRANSMEM 25261  
 FT TRANSMEM 0352C1D8BDAC24 CRC64;  
 SQ SEQUENCE 224 AA; 25261 MW; 0352C1D8BDAC24 CRC64;  
 QY 3 PAKLGTFTLMGKG 15  
 DB 202 FVALDGTLLSKG 214  
 Query Match 48.7%; Score 37; DB 1; Length 224;  
 Best Local Similarity 53.8%; Pred. No. 18;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

ID SPAP SALTY STANDARD; PRT; 224 AA.  
 AC P40700;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Surface presentation of antigens protein spap.  
 GN SPAP OR STM2890.  
 OS Salmonella typhimurium.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 NCBI\_TaxID=602;  
 RX MEDLINE=94008985; PubMed=8404849;  
 RA Groisman E.A., Ochman H.,  
 RT "Cognate gene clusters govern invasion of host epithelial cells by  
 RT Salmonella typhimurium and Shigella flexneri."  
 RL EMBO J. 12:3779-3787(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RC MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2."  
 RL Nature 413:852-856(2001).  
 CC -1- FUNCTION: INVOLVED IN A SECRETORY PATHWAY RESPONSIBLE FOR THE  
 CC SURFACE PRESENTATION OF DETERMINANTS NEEDED FOR THE ENTRY OF  
 CC SALMONELLA SPECIES INTO MAMMALIAN CELLS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE FLIP/MOPC/SPAP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X73525; CAA51925.1; -  
 DR EMBL: AE008832; AAL21770.1; -  
 DR PIR: S37308; S37308.  
 DR Styene; SG10468; spap.  
 DR InterPro: IPR002039; FLIP.  
 DR Pfam: PF00813; FLIP; 1.  
 DR PRINTS: PR01302; TYPE3IMPROT.  
 DR PRODOM: PD002586; FLIP; 1.  
 DR TIGRFAm: TIGR01102; yscr, 1.  
 DR PROSITE: PS01060; FLIP\_1; 1.  
 DR PROSITE: PS01061; FLIP\_2; 1.  
 KW Virulence; Transmembrane; Complete proteome.  
 FT TRANSMEM 8  
 FT TRANSMEM 55  
 FT TRANSMEM 161  
 FT TRANSMEM 185  
 FT TRANSMEM 205  
 FT TRANSMEM 224  
 FT TRANSMEM 25231  
 FT TRANSMEM 18241189C38D4F20 CRC64;  
 SQ SEQUENCE 224 AA; 25231 MW; 18241189C38D4F20 CRC64;  
 QY 3 PAKLGTFTLMGKG 15  
 DB 202 FVALDGTLLSKG 214  
 Query Match 48.7%; Score 37; DB 1; Length 224;  
 Best Local Similarity 53.8%; Pred. No. 18;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

RESULT 13  
 SPAP\_SALTY

RESULT 14  
 CYBH\_ALCEU

```

ID CYBH ALCEU STANDARD; PRT; 244 AA.
AC P31898;
DT 01-JUN-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable Ni/Fe-hydrogenase B-type cytochrome subunit.
GN HOXZ.
OS Alkaligenes eutrophus (Ralstonia eutropha).
OC Plasmid megaplasmid pHG1.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H16 / DSM 428 / ATCC 17699;
RX MEDLINE=93015670; PubMed=1383192;
RA Kortliueke C., Horstmann K., Schwartz E., Rohde M., Binsack R.,
RA Friedrich B.;
RT "A gene complex coding for the membrane-bound hydrogenase of
RT Alkaligenes eutrophus H16."
RL J. Bacteriol. 174:6277-6289(1992).
CC -1- FUNCTION: PROBABLE B-TYPE CYTOCHROME.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE HDPC/HVAC/HVNC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96433; AAA16463.1; -.
DR PIR; C43255; C43255.
DR InterPro: IPR000516; Ni_hydr_Cyrb.
DR Pfam; PF01292; Ni_hydr_Cyrb_1.
DR PRINTS; PR00161; N1HGNA5ECTYB.
DR PRODOM; PD005374; Ni_hydr_Cyrb_1.
DR PROSITE; PS00882; N1_HGNA5E_Cyrb_1; 1.
DR PROSITE; PS00883; N1_HGNA5E_Cyrb_2; 1.
KW Transmembrane; Electron transport; Heme; Plasmid.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 150 171 POTENTIAL.
FT TRANSMEM 204 221 POTENTIAL.
SQ SEQUENCE 244 AA; 27581 MW; 49435C5ECCT06961 CRC64;

Query Match 48.7%; Score 37; DB 1; Length 244;
Best Local Similarity 53.8%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 FAKLTGFTLMGKG 15
DB 162 FMILTGFAMVGG 174

RESULT 15
ID LPXA CHLMU STANDARD; PRT; 280 AA.
AC 09PUL1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase
DE (EC 2.3.1.129) (UDP-N-acetylglucosamine acyltransferase).
GN LPXA OR TC0818.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MOpn / N199;

```

```

RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uitterlind T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gilm M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF LIPID A, A
CC PHOSPHORYLATED GLYCOLIPID THAT ANCHORS THE LIPOLYSACCHARIDE TO
CC THE OUTER MEMBRANE OF THE CELL. (By similarity).
CC -1- CATALYTIC ACTIVITY: (R)-3-hydroxytetradecanoyl-[acyl-carrier
CC protein] + UDP-N-acetylglucosamine = [acyl-carrier protein] +
CC UDP-3-O-(3-hydroxytetradecanoyl)-N-acetylglucosamine.
CC -1- PATHWAY: Lipid A biosynthesis; first step.
CC -1- SUBUNIT: HOMOTRIMER (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRANSFERASE HEXAPEPTIDE REPEAT FAMILY.
CC LPXA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002347; AAF39620.1; -.
DR HSSP; P10440; 1LXA.
DR TIGR; TC0818; -.
DR InterPro: IPR001451; Hexapep_transf.
DR Pfam; PF00132; hexapep_8.
DR PROSITE; PS00101; HEXAPEP_TRANSFERSSES; 1.
KW Transferase; Acyltransferase; Lipid A biosynthesis; Lipid synthesis;
KW Repeat; Complete proteome.
SQ SEQUENCE 280 AA; 30672 MW; 4EE1B58937F0A63F CRC64;

Query Match 48.7%; Score 37; DB 1; Length 280;
Best Local Similarity 46.2%; Pred. No. 23;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 FAKLTGFTLMGKG 15
DB 43 YAVIDGFTTIGRG 55

```

Search completed: April 20, 2003, 13:07:41  
Job time : 3.92105 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 seconds  
(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524d-101

Perfect score: 76

Sequence: 1 LQFAKLTGFTLMGK 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: SP archaea:\*

2: SP bacteria:\*

3: SP fungi:\*

4: SP human:\*

5: SP invertebrate:\*

6: SP mammal:\*

7: SP mhc:\*

8: SP organelle:\*

9: SP phage:\*

10: SP plant:\*

11: SP rodent:\*

12: SP virus:\*

13: SP vertebrate:\*

14: SP unclassified:\*

15: SP virus:\*

16: SP bacteriap:\*

17: SP archaep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	63.2	507	10	Q9FY19 juniperus a
2	45	59.2	474	10	Q8W4S2
3	45	59.2	540	10	Q22699 arabidopsis
4	44	57.9	475	10	Q94EH4 arabidopsis
5	44	57.9	475	10	Q949Z1 arabidopsis
6	44	57.9	482	10	Q9LNG3 arabidopsis
7	43	56.6	194	16	Q9X7S4 arabidopsis
8	43	56.6	439	10	Q9SWS2 arabidopsis
9	43	56.6	443	10	Q9SWS3 arabidopsis
10	43	56.6	457	10	Q9ZUE7 arabidopsis
11	43	56.6	502	16	Q9RS41 arabidopsis
12	43	56.6	519	10	Q9LPS6 arabidopsis
13	43	56.6	1161	10	Q9LQD1 arabidopsis
14	41	53.9	360	16	Q9XZK6 streptomyces
15	41	53.9	442	10	Q9SDP3 allium cepa
16	41	53.9	468	10	Q9CAL5 arabidopsis

17	41	53.9	887	16	Q8XY49 ralstonia s
18	40	52.6	217	16	Q8YG50 bruceella me
19	40	52.6	355	10	Q94K16 pisum sativ
20	40	52.6	358	17	Q9YBY3 aeropyrum p
21	40	52.6	371	17	Q28833 archaeoglob
22	40	52.6	383	6	Q95IV2 macaca fasc
23	40	52.6	435	10	Q9LYJ5 arabidopsis
24	40	52.6	578	6	Q95JY8 macaca fasc
25	40	52.6	743	6	Q8WPJ3 macaca fasc
26	39	51.3	170	2	Q9X3N0 neisseria m
27	39	51.3	193	16	Q9CMN2 pasteurella
28	39	51.3	217	16	Q8YFT0 bruceella me
29	39	51.3	345	10	Q94CZ5 oryza sativ
30	39	51.3	359	6	Q9BGN2 oryctolagus
31	39	51.3	359	6	Q9BGN1 oryctolagus
32	39	51.3	380	10	Q94BV8 arabidopsis
33	39	51.3	382	10	Q95IM7 arabidopsis
34	39	51.3	425	10	Q8RYE4 arabidopsis
35	39	51.3	519	2	Q9RC14 oryza sativ
36	39	51.3	542	10	Q9SGY5 bacteroides
37	39	51.3	572	16	Q9JYH8 neisseria m
38	39	51.3	572	16	Q9JYH7 neisseria m
39	39	51.3	591	16	Q9J1L0 pseudomonas
40	39	51.3	689	10	Q9FPH8 arabidopsis
41	39	51.3	1020	16	Q9PB07 arabidopsis
42	39	51.3	1270	5	Q9GPN0 xylella fas
43	39	51.3	1348	5	Q9VDT3 caenorhabdi
44	38	50.0	244	2	Q9ANQ8 drosophila
45	38	50.0	312	10	Q9SUV3 bradyrhizob

## ALIGNMENTS

RESULT 1	ID	Q9FY19	PRELIMINARY;	PRT;	507 AA.
AC	Q9FY19	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Pollen major allergen 2 protein precursor.				
GN	JNA2.				
OS	Juniperus ashei (Ozark white cedar).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.				
OX	NCBI_TaxId=13101;				
RN	(1)				
RA	SEQUENCE FROM N.A.				
RC	TISSUE=MALE POLLEN;				
RX	MEDLINE=20403896; PubMed=10944464;				
RA	Yokoyama M., Miyahara M., Shimizu K., Kino K., Tsunoo H.;				
RT	"Purification, Identification and cDNA cloning of Jun a 2, the second				
RT	major allergen of mountain cedar pollen."				
RL	Biochem. Biophys. Res. Commun. 275:195-202(2000).				
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES				
CC	(POLYGLACTURONASES).				
EMBL	AF040653; CAC0582.1; ..				
DR	HSSP; P26509; IHRH.				
DR	InterPro; IPR000743; GH28.				
DR	InterPro; IPR00408; Reg_chrt condens.				
DR	Pfam; PF00295; Glyco_hydro_28; 1.				
DR	PROSITE; PS00502; POLYGLACTURONASE; UNKNOWN_1.				
DR	PROSITE; PS00626; RC01_2; UNKNOWN_1.				
KW	Cell wall; Glycosidase; Hydrolase; Signal.				
FT	SIGNAL 1				
FT	POTENTIAL.				
SEQUENCE	507 AA; 55730 MW; 282E0A5B958FE5A CRC64;				
Query Match	63.2%; Score 48; DB 10; Length 507;				
Best Local Similarity	60.0%; Pred. No. 3.1;				
Matches	9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;				
Qy	1 LQFAKLTGFTLMGK 15				

DB 141 MHPALTDENLWGTG 155

## RESULT 2

Q8W4S2 PRELIMINARY; PRT; 474 AA.  
 ID Q8W4S2  
 AC Q8W4S2  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE At1g60590/F8A5\_12  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY060568; AAJ31197.1; -.  
 DR InterPro; IPR000743; GH28.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN\_1.  
 DR SEQUENCE 474 AA; 51964 MW; 9631844B83E209C6 CRC64;  
 SQ  
 Query Match 59.2%; Score 45; DB 10; Length 474;  
 Best Local Similarity 46.7%; Pred. No. 9.9; Indels 0; Gaps 0;  
 Matches 7; Conservative 4; Mismatches 4;  
 QY 1 LQFAKLTGFTLMGKG 15  
 DB 140 IEFKLSGITIGNG 154  
 RESULT 3  
 ID Q22699 PRELIMINARY; PRT; 540 AA.  
 AC Q22699  
 DT 01-JUN-1998 (TREMBlrel. 05, Created)  
 DT 01-JUN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative polygalacturonase.  
 GN F8A5\_12.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Federespiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
 RA Lu M., Araujo R., Brendel V., Buehler E., Dewar K., Feng J., Kim C.,  
 RA Li Y., Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M.,  
 RA Vysotskaya V.S., Yu G., Ecker J., Theologis A., Davis R.W.;  
 RT "Genomic sequence of Arabidopsis."  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 CC EMBL; AC002292; AAB71972.1; -.  
 DR InterPro; IPR000743; GH28.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 DR Cell wall; Glycosidase; Hydrolase.  
 KM

SQ SEQUENCE 540 AA; 59631 MW; 532B79A848EBB904 CRC64;

Query Match 59.2%; Score 45; DB 10; Length 540;  
 Best Local Similarity 46.7%; Pred. No. 11;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15

DB 206 IEFKLSGITIGNG 220

## RESULT 4

Q94EH4 PRELIMINARY; PRT; 475 AA.  
 ID Q94EH4  
 AC Q94EH4  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE At1g48100/F21D18\_17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 CC EMBL; AF410319; AAK95305.1; -.  
 DR InterPro; IPR000743; GH28.  
 DR InterPro; IPR000408; Reg\_chromatins.  
 DR InterPro; IPR001412; RNA-synt\_1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_1; UNKNOWN\_1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN\_1.  
 DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.  
 KM Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 475 AA; 51380 MW; 4A42C4355827D30A CRC64;  
 Query Match 57.9%; Score 44; DB 10; Length 475;  
 Best Local Similarity 60.0%; Pred. No. 15;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 LQFAKLTGFTLMGKG 15  
 DB 164 LVFRLDGFPSGKG 178  
 RESULT 5  
 ID Q949Z1 PRELIMINARY; PRT; 475 AA.  
 AC Q949Z1  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative polygalacturonase PGL.  
 GN F21D18\_18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN

SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,  
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,  
 RA Carinini P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,  
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Tracy S.E.,  
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.,  
 RA "Full length cDNA of gene F1D18.18 (GI:8778527)."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 DR EMBL; AY050798; AK92733.1; -.  
 DR InterPro; IPR000743; GH28.  
 DR InterPro; IPR000408; Reg. chr. condens.  
 DR InterPro; IPR001412; tRNA-synt.1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_1; UNKNOWN\_1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN\_1.  
 DR Cell wall; Glycosidase; RC1.2; UNKNOWN\_1.  
 KW SEQUENCE 475 AA; 51352 MW; ACB066CF725F2C8F CRC64;

Query Match 57.9%; Score 44; DB 10; Length 475;  
 Best Local Similarity 60.0%; Pred. No. 15;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 DB 164 LVFYLRLDGFTEGSGK 178

## RESULT 6

Q9LNG3 PRELIMINARY; PRT; 492 AA.  
 AC Q9LNG3;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE F21D18.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;  
 RX SEQUENCE FROM N.A.  
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
 RA Shim P., Altafi H., Bei Q., Chin C., Chou J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RL "Genomic sequence for Arabidopsis thaliana BAC F21D18 from chromosome  
 1";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RX SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RX SEQUENCE FROM N.A.  
 RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.  
 RA Cheuk R., Shim P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chou J., Choi E.,  
 RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 DR EMBL; AC023673; AF79535.1; -.  
 DR InterPro; IPR000743; GH28.  
 DR InterPro; IPR000408; Reg. chr. condens.  
 DR InterPro; IPR001412; tRNA-synt.1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_1; UNKNOWN\_1.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 DR PROSITE; PS00626; RC1.2; UNKNOWN\_1.  
 KW Cell wall; Glycosidase; Hydrolyase.  
 SQ SEQUENCE 492 AA; 53395 MW; 42BAF98B9E30AC1A CRC64;

Query Match 57.9%; Score 44; DB 10; Length 492;  
 Best Local Similarity 60.0%; Pred. No. 16;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGKG 15  
 DB 164 LVFYLRLDGFTEGSGK 178

## RESULT 7

Q9X7S4 PRELIMINARY; PRT; 194 AA.  
 AC Q9X7S4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein SC07276.  
 GN SC07276 OR SC5H1.16.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1902;  
 RX SEQUENCE FROM N.A.  
 RA STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kiese H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,  
 RA Huang C.-H., Kiese H., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietztorrek A., Woodward J., Barrill B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RL "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL049863; CAB42941.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 194 AA; 20485 MW; D4BR04D35971F875 CRC64;

Query Match 56.6%; Score 43; DB 16; Length 194;  
 Best Local Similarity 63.6%; Pred. No. 9.3;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 KLTGFTLMGKG 15  
 DB 37 RITGFTLTGRG 47

## RESULT 8

Q9SWS2 PRELIMINARY; PRT; 439 AA.  
 AC Q9SWS2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Polyalacturonase PG2.  
 OS Glycine max (Soybean).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 ON NCBI\_TaxID=38477;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=99285288; PubMed=10356800;  
 RA Mahalingam R., Wang G., Knap H.T.;  
 RT "Polyalacturonase and polyalacturonase inhibitor protein: gene  
 isolation and transcription in Glycine max-Heterodera glycines  
 RT interaction.";  
 RL Mol. Plant Microbe Interact. 12:490-498 (1999).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYALACTURONASES).  
 CC EMBL: AF128267; AAD46484.1; -;  
 DR InterPro: IPR000743; GH28.  
 DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE: PS00502; POLYALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 439 AA; 47018 MW; D07AA96E3F42B7DB CRC64;

Query Match 56.6%; Score 43; DB 10; Length 439;  
 Best Local Similarity 53.3%; Pred. No. 21;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 LOPAKLTGFTLMGKG 15  
 Db 125 LEFSKLNKRTIRGKG 139

RESULT 9  
 O9SMS3 PRELIMINARY; PRT; 443 AA.  
 ID O9SMS3  
 AC O9SMS3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Polyalacturonase PG1.  
 OS Glycine max (Soybean).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 ON NCBI\_TaxID=38477;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=99285288; PubMed=10356800;  
 RA Mahalingam R., Wang G., Knap H.T.;  
 RT "Polyalacturonase and polyalacturonase inhibitor protein: gene  
 RT isolation and transcription in Glycine max-Heterodera glycines  
 RT interaction.";  
 RL Mol. Plant Microbe Interact. 12:490-498 (1999).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYALACTURONASES).  
 CC EMBL: AF128267; AAD46483.1; -;  
 DR InterPro: IPR000743; GH28.  
 DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE: PS00502; POLYALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 443 AA; 47335 MW; 13314187F6EBF33 CRC64;

Query Match 56.6%; Score 43; DB 10; Length 443;  
 Best Local Similarity 53.3%; Pred. No. 21;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 LOPAKLTGFTLMGKG 15  
 Db 129 LEFSKLNKRTIRGKG 143

RESULT 10  
 O9ZUE7

ID O9ZUE7 PRELIMINARY; PRT; 457 AA.  
 AC O9ZUE7;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE F508.2 protein.  
 GN F508.2  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=CV. COLUMBIA;  
 RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Liu S., Lenz C.,  
 RA Li J., Kremenetskaia I., Luros J., Altfeld H., Gonzalez A., Araujo R.,  
 RA Buehler E., Conn L., Conway A.B., Dunn P., Hansen N., Hutzler L.,  
 RA Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,  
 RA Becker J.R., Fedoripiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F508 sequence.";  
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=CV. COLUMBIA;  
 RA Theologis A.;  
 CC Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYALACTURONASES).  
 CC EMBL: AC005990; AAC98004.1; -;  
 DR InterPro: IPR000886; ER target.  
 DR InterPro: IPR000743; GH28.  
 DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN 1.  
 DR PROSITE: PS00502; POLYALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 457 AA; 49269 MW; 0EAA6551E566168 CRC64;

Query Match 56.6%; Score 43; DB 10; Length 457;  
 Best Local Similarity 53.3%; Pred. No. 22;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LOPAKLTGFTLMGKG 15  
 Db 154 LEFSKLNKRVFGKG 168

RESULT 11  
 O9RS41 PRELIMINARY; PRT; 502 AA.  
 ID O9RS41  
 AC O9RS41;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein DR2286.  
 GN DR2286.  
 OS Deinococcus radiodurans.  
 CC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 ON NCBI\_TaxID=1299;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.O., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Morfitt K.S., Qin H., Vang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1.";

RL Science 286:1571-1577(1999).  
 DR EMBL; AE002061; AAF11844.1; -  
 DR TIGR; DR2286; -  
 KW Hypothetical protein; Complete proteome  
 SQ SEQUENCE 502 AA; 53630 MW; 2E580505FE709AD0 CRC64;

Query Match 56.6%; Score 43; DB 16; Length 502;  
 Best Local Similarity 47.1%; Pred. No. 24;  
 Matches 8; Conservative 7; Mismatches 0; Indels 2; Gaps 1;  
 Oy 1 LOPAKTGF--TLMGK 15  
 Db 287 VERARLSGRVTLGRG 303

RESULT 12  
 Q9LPS6 PRELIMINARY; PRT; 519 AA.  
 AC Q9LPS6;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Putative cytochrome P450 (Cytochrome P450, putative).  
 GN F11F12.12 OR F17J6.8 OR ATG50560, F17J6.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NX NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,  
 RA Luros S., Schwartz J., Shim P., Toriumi M., Vayotskaia V.S.,  
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-NCV, COLUMBIA;  
 RC MEDLINE=21016719; PubMed=1130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federici N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldbylum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaylin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maletti R., Marshall A.,  
 RA Millecher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pat G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzenig S.L., Schwartz J.R., Shim P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utechtack T., Van Aken S., Vayenberg M., Vayotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RL "Sequence and analysis of chromosome I of the plant Arabidopsis  
 thaliana";  
 RT Nature 408:816-820(2000).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Caninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashitaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,  
 RA Suter T., Theologis A., Davis R.W.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL; AC079279; AAGS1202.1; -  
 DR EMBL; AY062573; AALJ2651.1; -

DR InterPro; IPR001128; Cytochrome\_P450.  
 DR InterPro; IPR001783; lsm\_binding.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 DR PROSITE; PS00693; LSM\_BINDING; UNKNOWN\_1.  
 KW Heme; Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 519 AA; 58472 MW; E7BD2048A7DE9D1A CRC64;

Query Match 56.6%; Score 43; DB 10; Length 519;  
 Best Local Similarity 61.5%; Pred. No. 25;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Oy 2 QPAKLTGFTLMGK 14  
 Db 239 QPSKLVGSLFGK 251

RESULT 13  
 Q9LQD1 PRELIMINARY; PRT; 1161 AA.  
 AC Q9LQD1;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE F28C11.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NX NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Shim P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
 RA Kim C., Altafi H., Bel O., Chin C., Chou J., Choi E., Conn L.,  
 RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vayenberg M., Yu G., Federici N.A., Theologis A.,  
 RA Ecker J.R.;  
 RL "Genomic sequence for Arabidopsis thaliana BAC F28C11 from chromosome  
 1";  
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
 RA Shim P., Altafi H., Bel B., Chin C., Chou J., Choi E., Conn L.,  
 RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu J., Liu S., Mukharzky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vayenberg M., Yu G., Davis R., Federici N., Theologis A.,  
 RA Ecker J.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shim P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharzky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vayenberg M., Yu G., Davis R., Federici N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTONAMASES).

```

DR EMBL; AC007945; AAF79584.1; -
DR InterPro; IPR000886; ER TARGET.
DR InterPro; IPR000743; GH28.
DR InterPro; IPR002885; PPR.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR Pfam; PF01535; PPR; 11.
DR TIGRFAMs; TIGR00756; PPR; 9.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE; PS00502; POLYGALACTONASE; 1.
DR Cell wall; Glycosidase; Hydrolase.
KW SEQUENCE. 1161 AA; 128197 MW; C67DCEFCBECAB99 CRC64;

Qy 1 LOPAKTGFMTMGKG 15
Db 827 LBSKLGKGVVFOCKG 841

Query Match 56.6%; Score 43; DB 10; Length 1161;
Best Local Similarity 53.3%; Pred. No. 56;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0.

RESULT 14
09KZK6
ID 09KZK6 PRELIMINARY; PRT; 360 AA.
AC 09KZK6;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative nucleotide phosphorylase.
GN SC03039 OR SCE34.20C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered contigs and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.R., Quail M.A., Kleiser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL353862; CAB88923.1; -
DR InterPro; IPR001451; Hexapep_transf.
DR InterPro; IPR001825; NTP_transferase.
DR Pfam; PF00132; Hexapep_3.
DR Pfam; PF00483; NTP_transferase; 1.

```

50	SEQUENCE	360 AA;	37642 MW;	9C50CDDFB2BCCF515 CRC64;	
	Query Match		53.9%;	Score 41;	DB 16; Length 360;
	Best Local Similarity	66.7%;		Pred. No. 39;	
	Matches	8; Conservative	3;	Mismatches	1; Indels 0; Gaps 0;
QY	4 AKLTGFTLMGK 15				
	:::-				
Db	266 AKLTGFTVVGEG 277				
	RESULT 15				
ID	Q9SDP3	PRELIMINARY;	PRT;	442 AA.	
AC	Q9SDP3;				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	APS-reductase.				
OS	Allium cepa (Onion).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	SpERMatoPhyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;				
OC	Allium.				
OX	NCBI_TaxID=4679;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. PUKEROHE LONG KEEPER; TISSUE=ROOT;				
RA	Pither-Joyce M.D., McCallum J.A.;				
RT	"Isolation of an APS-reductase cDNA clone from Allium cepa.";				
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF212155; AAF18999.1; -				
DR	InterPro; IPR004508; APS_reduc.				
DR	InterPro; IPR002500; PAPS_reduc.				
DR	InterPro; IPR000063; ThioRed.				
DR	Pfam; PF01507; PAPS_reduc; 1.				
DR	Pfam; PF00085; ThioRed; 1.				
DR	TIGRFAMs; TIGR00424; APS_reduc; 1.				
SQ	SEQUENCE 442 AA; 48710 MW; BE570C4F84A490CF CRC64;				
	Query Match		53.9%;	Score 41;	DB 10; Length 442;
	Best Local Similarity	50.0%;		Pred. No. 48;	
	Matches	7; Conservative	5;	Mismatches	2; Indels 0; Gaps 0;
QY	1 LOPAKLTGFTLMGK 14				
	:::-				
Db	127 LTPAKLTGPDVVK 140				

Search completed: April 20, 2003, 13:13:14  
Job time : 14.6711 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds  
(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524D-120

Perfect score: 79  
Sequence: 1 GIDIFASKNFHLQKN 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	514	2 S48730	Cry j II protein -
2	79	100.0	514	2 JC2498	second major aller
3	62	78.5	514	2 JC7100	polygalacturonase
4	54	54.4	507	2 JC7366	Jun a 2 protein -
5	43	54.4	419	2 T38199	conserved hypotet
6	42	53.2	156	2 D82888	hypothetical prote
7	42	53.2	305	2 T20685	hypothetical prote
8	41	53.2	357	1 Q0BER3	gene 58 protein -
9	41	51.9	259	2 T39029	hypothetical prote
10	41	51.9	630	2 S55645	transcription cont
11	40	50.6	598	2 B90589	sugar ABC transpor
12	40	50.6	1021	2 S64506	protein kinase BUB
13	39	49.4	64	2 T43075	hypothetical prote
14	39	49.4	280	2 T43075	Mg103 homolog KO4
15	39	49.4	280	2 D64211	hypothetical prote
16	39	49.4	349	2 S56488	hypothetical prote
17	39	49.4	353	2 A72410	choistamate mutase/
18	39	49.4	375	2 B96567	hypothetical prote
19	39	49.4	430	2 B9106	hypothetical prote
20	39	49.4	430	2 T30639	hypothetical prote
21	38	48.1	161	2 T50939	hypothetical prote
22	38	48.1	435	2 E86515	hypothetical prote
23	38	48.1	435	2 C72107	oligopeptide bindi
24	38	48.1	509	1 DEPGIP	peptide ABC transp
25	38	48.1	615	2 AD5502	ABC transporter d
26	38	48.1	625	2 G97076	alkaline phosphata
27	38	48.1	729	2 T45780	sugar transporter
28	38	48.1	824	2 T52835	H-NUC - human
29	38	48.1	1583	2 T14176	probable phosphati

30	37.5	47.5	1804	2 H96597	hypothetical prote
31	37	46.8	155	2 A31278	interleukin-2 prec
32	37	46.8	156	2 H97334	hypothetical prote
33	37	46.8	168	2 T08831	disease resistance
34	37	46.8	182	2 T46387	hypothetical prote
35	37	46.8	223	2 C86816	hypothetical prote
36	37	46.8	352	2 T42971	metalloproteinase
37	37	46.8	353	2 T42971	hypothetical prote
38	37	46.8	509	1 DEHUIP	diacylglycerol phosphatid
39	37	46.8	509	1 UC4241	probable membrane
40	37	46.8	517	2 S65305	probable glutathio
41	37	46.8	541	2 A81330	hypothetical prote
42	37	46.8	701	2 C84972	hypothetical prote
43	37	46.8	706	2 D90124	probable insectici
44	37	46.8	874	2 A10281	probable GRPase ac
45	37	46.8	1031	2 T38411	

#### ALIGNMENTS

##### RESULT 1

S48730 Cry j II protein - Japanese cedar

C:Species: Cryptomeria japonica (Japanese cedar)

C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999

C:Accession: S48730

R:Namba, M.; Kurose, M.; Torijoe, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.;

FEBS Lett. 353, 124-128, 1994

A:Title: Molecular cloning of the second major allergen, Cry j II, from Japanese cedar

A:Reference number: S48730; MUID:95010777; PMID:7926035

A:Accession: S48730

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-514 <NAM>

A:Cross-References: GB:D37765; MID:9577695; PIDN:BA07021.1; PID:di007598; PID:9577696

Query Match 100.0%; Score 79; DB 2; Length 514;

Best Local Similarity 100.0%; Pred. No. 2.9e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQKN 15  
DB 235 GIDIFASKNFHLQKN 249

##### RESULT 2

JC2498 second major allergen Cry j II precursor - Japanese cedar

C:Species: Cryptomeria japonica (Japanese cedar)

C:Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000

C:Accession: JC2498; PC2346; A60147

R:Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994

A:Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese

A:Reference number: JC2498; MUID:94271186; PMID:8002972

A:Accession: JC2498

A:Molecule type: mRNA

A:Residues: 1-514 <KOM>

A:Cross-References: DDBJ:D29772; MID:9506857; PIDN:BA06172.1; PID:9506858

A:Accession: PC2346

A:Molecule type: protein

A:Residues: 52-61 <KO2>

R:Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matunasi, T.

Allergy 45, 309-312, 1990

A:Title: Identification of the second major allergen of Japanese cedar pollen.

A:Reference number: A60147; MUID:90342988; PMID:2382797

A:Accession: A60147

A:Molecule type: protein

A:Residues: 55-64 <SAK>

C:Keywords: glycoprotein; pollen

F:1-54/Domain: signal sequence #status predicted <SIG>

F:55-460/Product: second major allergen Cry j #status predicted <MAT>

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 79; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 2.9e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQKN 15  
|||  
Db 235 GIDIFASKNFHLQKN 249

## RESULT 3

JC7100 polygalacturonase Cha o 2 - Japanese cypress

C/Species: Chamaecyparis obtusa (Japanese cypress)

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

C/Accession: JC7100; PC7026

R/Mori, T.; Yokoyama, M.; Komiyama, N.; Okano, M.; Kino, K.  
Biochem. Biophys. Res. Commun. 263, 166-171, 1999

A/Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A/Reference number: JC7100; PMID:10486272

A/Accession: JC7100

A/Molecule type: mRNA

A/Residues: 1-514 <MOR>

A/Accession: PC7026

A/Molecule type: protein

A/Residues: 51-62 <MO2>

Query Match 78.5%; Score 62; DB 2; Length 514;  
Best Local Similarity 78.6%; Pred. No. 0.0035;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQKN 14  
|||  
Db 235 GIDIFASKNFHLQKN 248

RESULT 4

JC7366 Jun a 2 protein - mountain cedar

C/Species: Juniperus ashei (mountain cedar)

C/Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C/Accession: JC7366; PC7093

R/Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.  
Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A/Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A/Reference number: JC7366

A/Accession: JC7366

A/Molecule type: mRNA

A/Residues: 1-507 <YOK>

A/Accession: PC7093

A/Molecule type: protein

A/Residues: 55-63 <YOK>

C/Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

C/Keywords: glycoprotein, pollen

Query Match 68.4%; Score 54; DB 2; Length 507;  
Best Local Similarity 71.4%; Pred. No. 0.097;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQKN 14  
|||  
Db 236 GIDIFASKNFHLQKN 249

RESULT 5

T38199 conserved hypothetical protein SPAC22F8.09 - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

R/Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Pohl, T.

submitted to the EMBL Data Library, August 1999

A/Reference number: Z21777

A/Accession: T38199

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-419 <LYN>

A/Cross-references: EMBL:AL109831; PIDN:CAB52719.1; GSPDB:GN00066; SPDB:SPAC22F8.09

A/Experimental source: strain 972h-; cosmid c22F8

A/Genetic: SPDB:SPAC22F8.09

A/Map position: 1

A/Introns: 46/2

Query Match 54.4%; Score 43; DB 2; Length 419;  
Best Local Similarity 63.6%; Pred. No. 7.6;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 FASKNFHLQKN 15  
|||  
Db 407 YSHKDFHLQKN 417

RESULT 6

D82888 hypothetical protein U949 [imported] - Ureaplasma urealyticum

C/Species: Ureaplasma urealyticum

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C/Accession: D82888

R/Glass, J.L.; Leikowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.

submitted to Genbank, February 2000

A/Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a

A/Reference number: A82870

A/Accession: D82888

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-156 <GLA>

A/Cross-references: GB:AE002142; GB:AF222894; NID:G6899447; PIDN:AAF30871.1; GSPDB:GN

A/Experimental source: serovar 3; biovar 1

C/Genetics:

A/Genetic code: SGC3

Query Match 53.2%; Score 42; DB 2; Length 156;  
Best Local Similarity 72.7%; Pred. No. 3.9;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 IDIFASKNFHL 12  
|||  
Db 66 IKIFKTNFHL 76

RESULT 7

T20685 hypothetical protein F10A3.12 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T20685

R/Lloyd, C.

submitted to the EMBL Data Library, March 1997

A/Reference number: Z19309

A/Accession: T20685

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-305 <WLL>

A/Cross-references: EMBL:Z92829; PIDN:CAB07348.1; GSPDB:GN00023; CESP:F10A3.12

A/Experimental source: clone F10A3

C/Genetics:

A/Genetic: CESP:F10A3.12

A/Map position: 5

A/Introns: 52/3; 82/1

Query Match 53.2%; Score 42; DB 2; Length 305;



Best Local Similarity 72.7%; Pred. No. 8.2;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 IDIFASKNFHL 12  
Db 112 IEIFAPKRFHL 122

# RESULT 8

Q08EP3  
gene 58 protein - saimiriine herpesvirus 1 (strain 11)

C/Species: saimiriine herpesvirus 1

A/Note: hoet Saimiri sciureus (common squirrel monkey)

C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 16-Jul-1999

C/Accession: B36812

R/Albrecht, J.

submitted to the EMBL Data Library, January 1992

A/Description: Primary structure of the herpesvirus saimiri genome.

A/Reference number: A36806

A/Accession: B36812

A/Molecule type: DNA

A/Residues: 1-357 <ALB>

A/Cross-references: GB:X64346; NID:960320; PIDN:CAA5681.1; PID:960379

A/Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W

A/Title: Primary structure of the herpesvirus saimiri genome.

A/Reference number: A37309; MUID:9233688; PMID:1321287

A/Contents: annotation; possible protein-coding frames

A/Note: neither amino acid nor nucleotide sequence is given

C/Genetics:

A/Status: preliminary

C/Superfamily: human herpesvirus 4 BMR2 protein

Query Match  
Best Local Similarity 53.2%; Score 42; DB 1; Length 357;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GIDIFASKNFH 11  
Db 171 GNGVPAKKNFH 181

# RESULT 9

T39029

hypothetical protein SPAC6C3.05 - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C/Accession: T39029

R/Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, February 1996

A/Reference number: Z21750

A/Accession: T39029

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-269 <DEV>

A/Cross-references: EMBL:Z69731; PIDN:CAR40280.1; GSPDB:GN00066; SPDB:SPAC6C3.05

A/Experimental source: strain 972h-; cosmid c6C3

C/Genetics:

A/Status: preliminary

A/Map position: 1

Query Match  
Best Local Similarity 51.9%; Score 41; DB 2; Length 269;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GIDIFASKNFHL 14  
Db 202 GLEISASKNFLEK 215

# RESULT 10

S55645  
transcription control protein 50 - equine herpesvirus 2

C/Species: equine herpesvirus 2

C/Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999

C/Accession: S55645

R/Reid, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A/Title: The DNA sequence of equine herpesvirus 2.

A/Reference number: S55594; MUID:95302501; PMID:7783207

A/Accession: S55645

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-630 <TEL>

A/Cross-references: GB:U20824; NID:9695172; PIDN:AAC13838.1; PID:9695223

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match  
Best Local Similarity 51.9%; Score 41; DB 2; Length 630;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 IDIFASKNFHL 15  
Db 614 LDIFSLNHLHRS 627

# RESULT 11

B90589

sugar ABC transporter permease protein [imported] - Mycoplasma pulmonis (strain UAB C

C/Species: Mycoplasma pulmonis

C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C/Accession: B90589

R/Chambaud, I.; Hellig, R.; Ferris, S.; Barde, V.; Samson, D.; Gallison, F.; Moszer, I

Nucleic Acids Res. 29, 2145-2153, 2001

A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p

A/Reference number: A99512; MUID:21267165; PMID:11353084

A/Accession: B90589

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-598 <KUR>

A/Cross-references: GB:AL445566; PID:91409003; PIDN:CAC13791.1; GSPDB:GN00153

A/Experimental source: strain UAB CTIP

C/Genetics:

A/Status: preliminary

A/Map position: 1

Query Match  
Best Local Similarity 50.6%; Score 40; DB 2; Length 598;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GIDIFASKNFH 11  
Db 383 GINIFASKNFH 393

# RESULT 12

S64506

protein kinase BUB1 (EC 2.7.1.-), checkpoint-associated - yeast (Saccharomyces cerevis

N/Alternate names: protein G7542; protein YGR188C

C/Species: Saccharomyces cerevisiae

C/Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 21-Jul-2000

C/Accession: S64506; A56354; S50224

R/Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.

submitted to the Protein Sequence Database, May 1996

A/Reference number: S64499

A/Accession: S64506

A/Molecule type: DNA

A/Residues: 1-1021 <ABR>

A/Status: preliminary

A/Experimental source: strain S288C

R/Roberts, B.T.; Farr, K.A.; Hoyt, M.A.

Mol. Cell. Biol. 14, 8282-8291, 1994

A/Title: The Saccharomyces cerevisiae checkpoint gene BUB1 encodes a novel protein kin

A/Reference number: A56354; MUID:95059057; PMID:7969164

A/Accession: A56354

A/Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-530, 'V', 532-1021 <ROB>  
 A:Cross-references: GB:I32027; NID:G475127; PID:AAA64894.1; PID:G475128  
 C:Genetics:  
 A:Gene: SGD:BUB1  
 A:Cross-references: SGD:S0003420; MIPS:YGR188C  
 A:Map position: 7R  
 C:Keywords: autophosphorylation; cell division control; phosphoprotein; phosphotransfera

Query Match  
 Best Local Similarity 50.6%; Score 40; DB 2; Length 1021;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DIFASKNFHLQ 15  
 DB 115 INLFNNFHESEN 128

RESULT 13  
 T43075  
 Hypothetical protein - Lactococcus lactis plasmid pMRC01  
 C:Species: Lactococcus lactis  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C:Accession: T43075  
 R:Dougherty, B.A.; Hill, C.; Weidman, J.F.; Richardson, D.R.; Venter, J.C.; Ross, R.P.  
 Mol. Microbiol. 29, 1029-1038, 1998  
 A>Title: Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid F  
 A:Reference number: Z22314  
 A:Accession: T43075  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-64 <DOU>  
 A:Cross-references: EMBL:AE001272; PID:AA656026.1  
 A:Experimental source: strain DPC3147  
 C:Genetics:  
 A:Genome: plasmid pMRC01  
 A>Note: ORF0002

Query Match  
 Best Local Similarity 49.4%; Score 39; DB 2; Length 64;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQ 13  
 DB 44 GILIFLKLNFHVQ 56

RESULT 14  
 S73916  
 MG103 homolog K04\_crf280 - Mycoplasma pneumoniae (strain ATCC 29342)  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C>Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 C:Accession: S73916  
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A>Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S73327; MID:97105885; PMID:8948633  
 A:Accession: S73916  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-280 <HIM>  
 A:Cross-references: EMBL:AE000058; GB:U00089; NID:G1674291; PID:AA696238.1; PID:G167429  
 A>Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Genetic code: SGC3

Query Match  
 Best Local Similarity 49.4%; Score 39; DB 2; Length 280;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DIFASKNFHLQ 13

DB 114 DLHSSNFHLQ 124

RESULT 15  
 D64211  
 Hypothetical protein MG103 - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 07-Dec-1999  
 C:Accession: D64211  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.  
 M.; Fierman, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,  
 C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A>Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MID:96026346; PMID:7569993  
 A:Accession: D64211  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-280 <TIGR>  
 A:Cross-references: GB:U39689; GB:I43967; NID:G1045773; PID:G1045781; TIGR:MG103  
 A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: SGC3

Query Match  
 Best Local Similarity 49.4%; Score 39; DB 2; Length 280;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DIFASKNFHLQ 13  
 DB 114 DLHSSNFHLQ 124

Search completed: April 20, 2003, 13:15:54  
 Job time : 8.07895 secs

GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-120

Perfect score: 79

Sequence: 1 GIDIFASKNFHLQKN 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	514	1	MPA2_CRYJA
2	43	54.4	419	1	YIY9_SCHPO
3	42	53.2	357	1	VG58_HSVSA
4	41	51.9	269	1	YD55_SCHPO
5	40	50.6	1021	1	BIB1_YEAST
6	39	49.4	273	1	BCHL_CHLAU
7	39	49.4	280	1	Y103_MYCCE
8	39	49.4	280	1	Y103_MYCCE
9	39	49.4	349	1	Y44F_SCHPO
10	39	49.4	437	1	PERG_LACDL
11	38	48.1	509	1	DLDH_PIG
12	38	48.1	824	1	CC27_HUMAN
13	37	46.8	155	1	IL2_RAT
14	37	46.8	223	1	PURQ_LACIA
15	37	46.8	352	1	SMP_SERMA
16	37	46.8	509	1	DLDH_HUMAN
17	37	46.8	509	1	DLDH_HUMAN
18	37	46.8	701	1	Y363_BUCAI
19	37	46.8	1031	1	YD69_SCHPO
20	36	45.6	137	1	MESC_LETME
21	36	45.6	153	1	IL2_HUMAN
22	36	45.6	154	1	IL2_CERTO
23	36	45.6	154	1	IL2_MACNU
24	36	45.6	208	1	PSB2_DROME
25	36	45.6	234	1	A29B_DROME
26	36	45.6	284	1	YMX7_YEAST
27	36	45.6	330	1	UL16_HSVSA
28	36	45.6	331	1	GPDA_CLOPE
29	36	45.6	358	1	Y39L_YEAST
30	36	45.6	412	1	FTS2_ENTFA
31	36	45.6	620	1	Y618_YEAST
32	36	45.6	672	1	MIS6_SCHPO
33	36	45.6	706	1	HDAL_YEAST

34	36	45.6	752	1	CO2_HUMAN	P06681	homo sapien
35	36	45.6	768	1	YB23_HUMAN	O91177	homo sapien
36	36	45.6	768	1	EXG1_COCCA	P49426	cochliobol
37	35.5	44.9	342	1	QWEA_STREP	O97915	streptococc
38	35.5	44.9	342	1	QWEA_STREP	O97915	streptococc
39	35	44.3	132	1	TVAA_MOUSE	P06323	mus musculu
40	35	44.3	140	1	NDK_RHOSU	P95653	rhodovulum
41	35	44.3	147	1	YBAK_BACSU	P50862	bacillus su
42	35	44.3	155	1	VG06_VACC	P21027	vaccinia vi
43	35	44.3	165	1	VG06_VACC	P32996	variola vir
44	35	44.3	170	1	YHE3_YEAST	P38728	saccharomyc
45	35	44.3	194	1	YCEF_ECO57	P58626	escherichia

## ALIGNMENTS

RESULT 1  
MPA2\_CRYJA STANDARD; PRT; 514 AA.  
ID MPA2\_CRYJA  
AC P43212;  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE Possible polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)  
OS (Major pollen allergen Cry j 2) (Cry j II).  
DE Cryptomeria japonica (Japanese cedar).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.  
OX NCBI\_TaxId=3369;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Pollen;  
RX MEDLINE=95010777; PubMed=7926035;  
RA Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.,  
RT Usui M., Kurimoto M.,  
RT "Molecular cloning of the second major allergen, Cry j II, from  
RT Japanese cedar pollen.";  
RL FEBS Lett. 353:124-128(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pollen;  
RX MEDLINE=94271186; PubMed=8002972;  
RA Komiyama N., Sone T., Shimizu K., Morikubo K., Kino K.,  
RT "cDNA cloning and expression of Cry j II the second major allergen of  
RT Japanese cedar pollen.";  
RL Biochem. Biophys. Res. Commun. 201:1021-1028(1994).  
RN [3]  
RP SEQUENCE OF 55-64.  
RX MEDLINE=90342988; PubMed=2382797;  
RA Sakaguchi M., Inoue S., Tanai M., Ando S., Usui M., Matubasi T.,  
RT "Identification of the second major allergen of Japanese cedar  
RT pollen.";  
RL Allergy 45:309-312(1990).  
CC -I- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-  
CC galactosiduronic linkages in pectate and other galacturonans.  
CC -I- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).  
CC -I- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
CC (POLYGALACTURONASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: D37765; BAA07021.1; -;  
CC EMBL: D29772; BAA06172.1; -;  
CC HSSP: P26509; 1BHE.  
CC InterPro: IPR000743; GH28.  
CC Pfam: PF00295; Glyco\_hydro\_28; 1.

DR PROSITE; PS00502; POLYGALACTURONASE, 1.  
 KM Hydroxylase; Glycosidase; Cell wall; Signal; Zymogen; Fruit ripening;  
 KW Amyloplase; Glycoprotein; Allergen.  
 FT SIGNAL 1 22  
 FT PROPER 23 45  
 FT CHAIN 46 433  
 FT PROPE 434 514  
 FT ACT SITE 278 278  
 FT CARBOHYD 460 460  
 FT CARBOHYD 472 472  
 FT CONFLICT 5 5  
 FT CONFLICT 12 12  
 FT CONFLICT 34 35  
 FT CONFLICT 37 37  
 FT CONFLICT 88 88  
 FT CONFLICT 98 98  
 FT CONFLICT 451 451  
 FT CONFLICT 454 454  
 FT CONFLICT 504 504  
 FT CONFLICT 507 507  
 SQ SEQUENCE 514 AA; 56645 MW; 624611C3FA8D6302 CRC64;

Query Match 100.0%; Score 79; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQKN 15  
 DB 235 GIDIFASKNFHLQKN 249

RESULT 2  
 YTY9\_SCHPO STANDARD; PRT; 419 AA.  
 AC Q9UT14;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C22F8.09 in chromosome 1.  
 GN SPAC22F8.09.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 CC NCBI\_Taxid=4896;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Miblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkert G., Aert R., Robben J., Grympeiz B.,  
 RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs W., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehmach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Driano S., Gloux S., Lelaure V., Motier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huest S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento G.,  
 RA Dominguez A., Revuelta J.U., Moreno S., Armstrong J., Folsburg S.L.,  
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC -1- SIMILARITY: BELONGS TO THE GLTSCR2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL109831; CAB52719.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 419 AA; 48129 MW; 9C2A72B639E61DAA CRC64;

Query Match 54.4%; Score 43; DB 1; Length 419;  
 Best Local Similarity 63.6%; Pred. No. 3.1;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 FASKNFHLQKN 15  
 DB 407 YSHKDFHLQKN 417

RESULT 3  
 VG58\_HSVSA STANDARD; PRT; 357 AA.  
 AC 001053;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE Gene 58 protein.  
 GN 58 OR EHEF5.  
 OS Herpesvirus saimiri (strain 11).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Rhadinovirus.  
 CC NCBI\_Taxid=10383;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=92333688; PubMed=1321287;  
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
 RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
 RA Honess R.W.;  
 RT "Primary structure of the herpesvirus saimiri genome."  
 RL J. Virol. 66:5047-5058(1992).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230228; PubMed=1314457;  
 RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
 RT "Analysis of nucleotide sequence of the rightmost 43 kbp of  
 RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic  
 RT organization between HVS and Epstein-Barr virus."  
 RL Virology 188:296-310(1992).  
 CC -1- SIMILARITY: TO EBV BMRF2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X64346; CAA45681.1; -  
 DR EMBL; M86409; AAA46134.1; -  
 DR PIR; B36812; Q0BEP3.  
 SQ SEQUENCE 357 AA; 40526 MW; 390F01A23F9F56B1 CRC64;

Query Match 53.2%; Score 42; DB 1; Length 357;  
 Best Local Similarity 63.6%; Pred. No. 4;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GIDIFASKNFH 11  
Db 171 GNGVFLKNFH 181

## RESULT 4

YD55\_SCHPO STANDARD: PRT; 269 AA.  
AC Q10308;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein C6C3.05 in chromosome I.  
GN SPAC6C3.05.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Brooks K., Peat N., Hayles J., Baker S., Basham D., Bowman S., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean U., Mooney P., Moulis S., Mungall K., Murphy L., Nidlett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E., Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S., Skelton G., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Voicakeet G., Aert R., Robben J., Grymoprez B., Wellens I., Vanstreel E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Tessey D., Barrett B.G., Nurse P.;  
RA "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z69731; CAB40280.1; -  
DR Hypothetical protein.  
KW  
SQ SEQUENCE 269 AA; 31234 MW; 0C3DF87A138CA5BE CRC64;

Query Match 51.9%; Score 41; DB 1; Length 269;  
Best Local Similarity 57.1%; Pred. NO. 4.4;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 GIDIFASKNFH 14  
Db 202 GIDIFASKNFH 215

## RESULT 5

BUB1\_YEAST

ID BUB1\_YEAST STANDARD: PRT; 1021 AA.  
AC P41695;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Checkpoint serine/threonine-protein kinase BUB1 (EC 2.7.1.1-).  
GN BUB1 OR YGR188C OR G7542.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9288C;  
RX MEDLINE=95059057; PubMed=7969164;  
RA Roberts B.T., Farr K.A., Hoyt M.A.;  
RT "The Saccharomyces cerevisiae checkpoint gene BUB1 encodes a novel protein kinase.";  
RL Mol. Cell. Biol. 14:8282-8291(1994).

RP SEQUENCE FROM N.A.  
RC STRAIN=9288C;  
RX MEDLINE=97279231; PubMed=9133739;  
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M., Nombela C.;  
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm of Saccharomyces cerevisiae chromosome VII.";  
RL Yeast 13:357-363(1997).

CC -1- FUNCTION: INVOLVED IN CELL CYCLE CHECKPOINT ENFORCEMENT. CATALYZES THE PHOSPHORYLATION OF BUB3 AND ITS AUTOPOHOSPHORYLATION.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- PTM: AUTOPOHOSPHORYLATED.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC BUB1 SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 1 CD1 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L32027; AAA64894.1; -  
DR EMBL; Z72973; CAA97214.1; -  
DR EMBL; X99074; CAA67524.1; -  
DR SGD; S0003420; BUB1.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR Prodom: PPO00001; Euk\_pkinase; 1.  
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
KW Transferrase; Serine/threonine-protein kinase; ATP-binding;  
KW Cell cycle; Phosphorylation; Nuclear protein.  
FT DOMAIN 55 304  
FT NP\_BIND 705 1021 PROTEIN KINASE.  
FT BINDING 711 719 ATP (BY SIMILARITY).  
FT ACT\_SITE 833 833 ATP.  
FT MUTAGEN 733 733 BY SIMILARITY.  
FT CONFLICT 531 531 K->R; LOSS OF ACTIVITY.  
SQ SEQUENCE 1021 AA; 117868 MW; 6D76FC960775D3F9 CRC64;

Query Match 50.6%; Score 40; DB 1; Length 1021;  
Best Local Similarity 50.0%; Pred. NO. 28;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 IDIFASKNFH 15  
Db 115 INFLSNPFHESN 128

```

CC Bacteria;Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
CC NCBI_TaxId=2097;
RC RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7565993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann J.R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Meldrum J.F., Small K.V., Sandusky M., Fultmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Inoué T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN RN [2]
RP SEQUENCE OF 86-208 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing";
RL J. Bacteriol. 175:7918-7930(1993).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U39690; AAC71321.1; -
DR EMBL; U02170; AAD12452.1; -.
DR TIGR; MG103; -.
DR InterPro; IPR003802; DUF199.
DR Pfam; PF02650; DUF199; 1.
DR TRIPFAMS; TIGR00647; MG103; 1.
DR Hypothetical protein; Complete proteome.
RT CONFLICT 86 L -> F (IN REF. 2).
SQ SEQUENCE 280 AA; 32864 MW; 022BC4D1D735ARD CRC64;
-----
QY 3 DIFASKNFHLQ 13
|:|:|||||
Db 114 DLINSNPFHLQ 124

Query Match 49.4%; Score 39; DB 1; Length 280;
Best Local Similarity 63.6%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DIFASKNFHLQ 13
|:|:|||||
Db 114 DLINSNPFHLQ 124

RESULT 8
Y103_MYCPN STANDARD; PRT; 280 AA.
ID Y103_MYCPN
AC P75530;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG103 homolog (K04_orf280).
GN MPN241 OR MP590.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2104;
RN RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105085; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plogens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE000058; AAB96238.1; -  
 DR InterPro; IPR003802; DUF199.  
 DR Pfam; PF02650; DUF199; 1.  
 DR TIGRPFAMs; TIGR00647; MG103; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 280 AA; 32194 MW; 5842D83CBFC26AAA CRC64;

Query Match 49.4%; Score 39; DB 1; Length 280;  
 Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DIFASKNFHLQ 13  
 Db 114 DLHSSNPHLQ 124

RESULT 9  
 YAAF SCHPO STANDARD; PRT; 349 AA.  
 AC Q09732;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C31A2.15c in chromosome 1.  
 GN SPAC31A2.15c.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 RN NCB1\_TaxID=4896;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brookes K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skellon J., Stimmings M., Squares R., Squares K., Stevens K.,  
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whithread S.,  
 RA Woodward J., Volkart E., Aert R., Robben J., Grymporter B.,  
 RA Wellens I., Vanstreels B., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Punnett B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Mortier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.U., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z50113; CA90473.1; -  
 DR Hypothetical protein.  
 KM SEQUENCE 349 AA; 40938 MW; 3143B4EB763D280A CRC64;

Query Match 49.4%; Score 39; DB 1; Length 349;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 IDIFASKNFHLQKN 15  
 Db 143 VDSFLOKNFPMVKN 156

RESULT 10  
 PEPG LACDL STANDARD; PRT; 437 AA.  
 AC P94869;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aminopeptidase G (EC 3.4.22.-).  
 GN PEPG.  
 OS Lactobacillus delbrueckii (subsp. lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 RN NCB1\_TaxID=29397;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 7290;  
 RX MEDLINE=97195798; PubMed=9043129;  
 RA Klein J.R., Schick J., Heinrich B., Plapp R.,  
 RA "Lactobacillus delbrueckii subsp. lactis DSM7290 pepg gene encodes a  
 RT novel cysteine aminopeptidase.";  
 RL Microbiology 143:527-537(1997).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z71782; CA96465.1; -  
 DR HESP; O13867; 2CR5.  
 DR InterPro; IPR004134; Pept C1 like.  
 DR InterPro; IPR000668; Peptidase\_C1.  
 DR InterPro; IPR000169; SHprotease.  
 DR Pfam; PF03051; Pept C1-like; 1.  
 DR PRINTS; PR00705; PAPA1N.  
 DR PROSITE; PS00639; THOL\_PROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THOL\_PROTEASE\_HIS; 1.  
 DR PROSITE; PS00640; THOL\_PROTEASE\_ASN; FALSE-NEG.  
 KM Hydrolyse; Thiol protease; Aminopeptidase.  
 FT ACT SITE 70  
 FT ACT SITE 361  
 FT ACT SITE 382  
 FT ACT SITE 382  
 SQ SEQUENCE 437 AA; 49705 MW; 2099C62E57181FC5 CRC64;

Query Match 49.4%; Score 39; DB 1; Length 437;  
 Best Local Similarity 75.0%; Pred. No. 17;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 8 KNFHLQKN 15  
 Db 234 KNFHLQKN 241

RESULT 11  
 ID DLDPH PIG STANDARD; PRT; 509 AA.  
 AC P09623;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Dihydrolipoamide dehydrogenase, mitochondrial precursor (EC 1.8.1.4).  
 GN DLDP OR LAD.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88087005; PubMed=3693355;  
 RA Ostrowski G., Robinson B.H.;  
 RT "Isolation and sequence determination of cDNA clones for porcine and  
 human lipoprotein dehydrogenase. Homology to other disulfide  
 oxidoreductases." J. Biol. Chem. 262:17313-17318(1987).  
 RL J. Biol. Chem. 262:17313-17318(1987).  
 CC -1- FUNCTION: LIPOAMIDE DEHYDROGENASE IS A COMPONENT OF THE GLYCINE  
 CLEAVAGE SYSTEM AS WELL AS OF THE ALPHA-KETOACID DEHYDROGENASE  
 COMPLEXES.  
 CC -1- CATALYTIC ACTIVITY: Dihydrolipoamide + NAD(+) = Lipoamide + NADH.  
 CC -1- COFACTOR: FAD.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.  
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE  
 OXIDOREDUCTASES CLASS-1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; J03489; AAA31069.1; -  
 DR PIR; A28448; DEPGP.  
 DR HSSP; P18925; 3IAD.  
 DR InterPro; IPR001137; FAD pyr redox.  
 DR InterPro; IPR001100; Pyr\_redox.  
 DR Pfam; PF00070; Pyr\_redox; 1.  
 DR Pfam; PF02852; Pyr\_redox; 1.  
 DR PRINTS; PR00368; FADPYR.  
 DR PRINTS; PR00411; PNDPDTASE1.  
 DR PRODOM; PD000139; FAD pyr redox; 1.  
 DR PROSITE; PS00076; PYRIDINE-REDOX\_1; 1.  
 KW Redox-active center; Oxidoreductase; NAD; Flavoprotein; FAD;  
 KW Mitochondrion; Transil peptide.  
 KM TRANSIT 1  
 FT CHAIN 35 509 MITOCHONDRION.  
 FT NP\_BIND 43 73 DIHYDROLIPOAMIDE DEHYDROGENASE.  
 FT DISULFID 80 85 FAD (ADP PART) (PROBABLE).  
 FT NP\_BIND 345 355 REDOX-ACTIVE.  
 FT ACT\_SITE 487 487 FAD (FLAVIN PART) (BY SIMILARITY).  
 SQ SEQUENCE 509 AA; 54185 MW; 38A0469FED071300 CRC64;

Query Match 48.1%; Score 38; DB 1; Length 509;  
 Best Local Similarity 62.5%; Pred. No. 30;  
 Matches 10; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 1 GIDIFASKNPH--LQK 14  
 Db 252 GIDMEVSKNFORLQK 267

RESULT 12

CC27 HUMAN  
 ID CC27 HUMAN STANDARD; PRT; 824 AA.  
 AC P30250; O16349; Q96F35;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein CDC27H (Cell division cycle protein 27 homolog) (H-NUC).  
 GN CDC27.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94052097; PubMed=8234252;  
 RA Tygdenreich S., Boguski M.S., Seldin M., Hieter P.A.;  
 RT "Linking yeast genetics to mammalian genomes: identification and  
 mapping of the human homolog of CDC27 via the expressed sequence tag  
 (EST) data base." Proc. Natl. Acad. Sci. U.S.A. 90:10031-10035(1993).  
 RL Cell Growth Differ. 6:199-210(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95275739; PubMed=7756179;  
 RA Chen P.L., Ueng Y.C., Durfee T., Chen K.C., Yang-Feng T., Lee W.H.;  
 RT "Identification of a human homologue of yeast nuc2 which interacts  
 with the retinoblastoma protein in a specific manner." J.  
 Cell Growth Differ. 6:199-210(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Uterus;  
 RA Strausberg R.;  
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RL -1- SUBUNIT: INTERACTS WITH RB.  
 CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC -1- SIMILARITY: BELONGS TO THE CDC27/NUC2 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U00001; AAA60471.1; -  
 DR EMBL; S78234; AAB34378.1; -  
 DR EMBL; BC011656; AAB11656.1; -  
 DR Genew; HGNC:1728; CDC27.  
 DR MIM; 116946; -  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 8.  
 DR SMART; SMO0028; TPR; 7.  
 KW Repeat; TPR repeat; Nuclear protein.  
 KM Repeat: TPR repeat: Nuclear protein.  
 FT REPEAT 84 114 TPR 1.  
 FT REPEAT 115 148 TPR 2.  
 FT REPEAT 499 532 TPR 3.  
 FT REPEAT 567 600 TPR 4.  
 FT REPEAT 602 634 TPR 5.  
 FT REPEAT 635 668 TPR 6.  
 FT REPEAT 670 702 TPR 7.  
 FT REPEAT 704 736 TPR 8.  
 FT REPEAT 737 770 TPR 9.  
 FT CONFLICT 319 403 K -> KTERVQ (IN REF. 3).  
 FT CONFLICT 403 403 K -> E (IN REF. 3).  
 FT CONFLICT 460 460 MISSING (IN REF. 1).  
 FT CONFLICT 715 715 A -> R (IN REF. 1).  
 SQ SEQUENCE 824 AA; 91867 MW; E6C8F59C1EFDCBA CRC64;

Query Match 48.1%; Score 38; DB 1; Length 824;  
 Best Local Similarity 40.0%; Pred. No. 51;  
 Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;



OY 1 GIDIFASKNFHLQKN 15  
 DB 535 GMEIYSTLMLHOKD 549

## RESULT 13

IL2\_RAT  
 ID IL2\_RAT STANDARD; PRT; 155 AA.  
 AC P17108;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).  
 GN IL2 OR IL-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_Taxid=10116;  
 [1]

SEQUENCE FROM N.A.  
 RX MEDLINE=89339608; Pubmed=2788130;  
 RA McKnight A.J., Mason D.W., Barclay A.N.;  
 RT "Sequence of rat interleukin 2 and anomalous binding of a mouse  
 RT interleukin 2 cDNA probe to rat MHC class II-associated invariant  
 RT chain mRNA".  
 RL Immunogenetics 30:145-147(1989).  
 CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR  
 CC MYTOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL  
 CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE  
 CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-  
 CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M22899; AAA41427.1; -  
 DR PIR; A31278; A31278.  
 DR PIR; A45882; A45882.  
 DR HSSP; P01585; SINK.  
 DR InterPro; IPR000779; Interleukin-2.  
 DR Pfam; PF00715; IL2; 1.  
 DR PRINTS; PR00265; INTERLEUKIN2.  
 DR ProDom; PD003649; Interleukin-2; 1.  
 DR SMART; SM00189; IL2; 1.  
 DR PROSITE; PS00424; INTERLEUKIN\_2; 1.  
 KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;  
 KW T-cell.  
 FT SIGNAL 1 20 BY SIMILARITY.  
 FT CHAIN 21 155 INTERLEUKIN-2.  
 FT CARBOHYD 23 23 O-LINKED (GALNAAC... ) (BY SIMILARITY).  
 FT DISUPTID 78 126 BY SIMILARITY.  
 SQ SEQUENCE 155 AA; 17632 MW; 67A8554A73BF30A0 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 155;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 IDIFASKNFHLQ 13  
 DB 90 LDLTQSKSFHLE 101

RESULT 14  
 ID PURO\_LACIA STANDARD; PRT; 223 AA.  
 AC Q9CEB7;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phosphoribosyltransferase 1 (EC 6.3.5.3) (PGAM  
 DE synthase I).  
 GN PURO OR IL1531.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 ON NCBI\_Taxid=1360;  
 [1]

SEQUENCE FROM N.A.  
 RX STRAIN=IL1403;  
 RX MEDLINE=21235186; Pubmed=11337471;  
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403.".  
 RL Genome Res. 11:731-753(2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-  
 CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-  
 CC (formamido)-N(1)-(5-phospho-D-ribosyl)acetamide + L-glutamate.  
 CC -1- PATHWAY: De novo purine biosynthesis; fourth step.  
 CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURO AND PURL.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE006384; AK05629.1; -  
 DR InterPro; IPR000991; GATASE\_1.  
 DR InterPro; IPR002818; ThJd.  
 DR Pfam; PF01965; ThJd; 1.  
 DR PROSITE; PS00442; GATASE\_TYPE\_I; FALSE NEG.  
 KW Purine biosynthesis; Ligase; Glutamine amidotransferase;  
 KW Complete proteome.  
 FT ACT\_SITE 86 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 223 AA; 24205 MW; A8FED3AE08A1262 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 223;  
 Best Local Similarity 66.7%; Pred. No. 19;  
 Matches 8; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

OY 1 GIDIFAS-KNF 10  
 DB 209 GVDLFASVLEKNF 220

RESULT 15  
 SMP\_SERMA  
 ID SMP\_SERMA STANDARD; PRT; 352 AA.  
 AC Q06517;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Extracellular minor metalloprotease precursor (EC 3.4.24.-).  
 GN SMP.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 ON NCBI\_Taxid=615;  
 [1]

SEQUENCE FROM N.A., AND SEQUENCE OF 53-57.  
 RC STRAIN=ATCC 21074;  
 RX MEDLINE=93194081; Pubmed=8449415;  
 RA Kwon Y.T., Lee H.H., Rho H.M.;  
 RT "Cloning, sequencing, and expression of a minor protease-encoding  
 RT gene from Serratia marcescens ATCC21074.";

```

RL Gene 125:75-80(1993).
CC -1- COFACTOR: BINDS 1 ZINC ION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE OPTIMAL TEMPERATURE AND PH FOR ENZYME ACTIVITY
CC ARE 50 DEGREES CELSIUS AND PH 8.0.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M59854; AAA26553.1; -.
CC PIR: J0153; J0153.
CC HSSP: P05806; INPC.
DR MEROPS: M04.UFW; -.
DR InterPro: IPR001570; Peptidase_M4.
DR InterPro: IPR000130; Zn_MTPeptidase.
DR Pfam: PF02868; Peptidase_M4_C; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KM Hydrolase; Metalloprotease; zinc; Signal; Zymogen.
FT SIGNAL 1
FT PROPEP 52
FT CHAIN 53 352 ACTIVATION PEPTIDE.
FT METAL 160 160 EXTRACELLULAR MINOR METALLOPROTEASE.
FT ACT_SITE 161 161 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 164 164 BY SIMILARITY.
FT METAL 164 164 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 184 184 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 352 AA; 38515 MW; D10IC956426B163C CRC64;

```

```

Query Match 46.8%; Score 37; DB 1; Length 352;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 8; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

```

```

QY 3 DIFAS--KNFHLOK 14
DB 188 DVGSLVKQFHLOQ 201

```

Search completed: April 20, 2003, 13:07:43  
 Job time : 4.92105 secs

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 seconds  
(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524d-120  
Perfect score: 79  
Sequence: 1 GIDIFASKNFHLQKN 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	68.4	507	10 Q9FY19	Q9FY19 Juniperus a
2	45	57.0	503	13 Q9W6Y2	Q9W6Y2 oncorhynch
3	43	54.4	252	2 Q9AHX1	Q9AHX1 carsonella
4	43	54.4	1360	5 O15807	O15807 cryptospori
5	43	54.4	1431	5 Q9XVH6	Q9XVH6 cryptospori
6	42	53.2	156	16 Q9PQ31	Q9PQ31 ureaplasma
7	42	53.2	305	5 O45343	O45343 caenorhadi
8	41	51.9	630	12 O66652	O66652 equine heip
9	40	50.6	251	10 O81407	O81407 sinapis arv
10	40	50.6	445	16 O8XOS9	O8XOS9 clostridium
11	40	50.6	598	16 Q98PV1	Q98PV1 mycoplasma
12	40	50.6	859	5 Q9VR4	Q9VR4 mycoplasma
13	39.5	50.0	490	10 Q9XEV7	Q9XEV7 oryza sativ
14	39	49.4	64	2 O87205	O87205 lactococcus
15	39	49.4	122	12 O11321	O11321 molluscum c
16	39	49.4	204	16 Q8R5N9	Q8R5N9 fusobacteri

17	39	49.4	335	4 Q9Y382	Q9Y382 homo sapien
18	39	49.4	353	16 Q9WY02	Q9WY02 thermotoga
19	39	49.4	375	10 Q9SSR8	Q9SSR8 arabidopsis
20	39	49.4	401	10 O24554	O24554 zinnia eleg
21	39	49.4	430	10 Q9AW73	Q9AW73 quillardia
22	39	49.4	565	12 Q98205	Q98205 molluscum c
23	39	49.4	623	16 Q8R669	Q8R669 fusobacteri
24	39	49.4	904	5 Q9U011	Q9U011 plasmodium
25	39	49.4	1073	5 Q9G0N3	Q9G0N3 caenorhadi
26	39	48.1	34	5 Q8T3N5	Q8T3N5 dreosiphila
27	38	48.1	161	3 Q9P3N1	Q9P3N1 neurospora
28	38	48.1	170	2 O32478	O32478 rhodobacter
29	38	48.1	248	2 Q93UB6	Q93UB6 carsonella
30	38	48.1	248	2 Q93UB9	Q93UB9 carsonella
31	38	48.1	253	2 Q93UB6	Q93UB6 carsonella
32	38	48.1	258	5 Q9VR67	Q9VR67 dreosiphila
33	38	48.1	324	4 Q9Y310	Q9Y310 homo sapien
34	38	48.1	348	12 O57103	O57103 monkeypox v
35	38	48.1	348	12 O57108	O57108 monkeypox v
36	38	48.1	348	12 O57277	O57277 monkeypox v
37	38	48.1	349	12 O57099	O57099 monkeypox v
38	38	48.1	349	12 O57100	O57100 monkeypox v
39	38	48.1	349	12 O57101	O57101 monkeypox v
40	38	48.1	349	12 O57102	O57102 monkeypox v
41	38	48.1	349	12 O57291	O57291 monkeypox v
42	38	48.1	435	16 Q9V5X2	Q9V5X2 chlamydia p
43	38	48.1	506	5 Q9V5X2	Q9V5X2 dreosiphila
44	38	48.1	615	16 Q8YK05	Q8YK05 anabaena sp
45	38	48.1	625	16 Q97J52	Q97J52 clostridium

## ALIGNMENTS

RESULT 1	ID	Q9FY19	PRELIMINARY;	PRT;	507 AA.
AC	Q9FY19	01-MAR-2001 (T-EMBLrel. 16, Created)			
DT	01-MAR-2001 (T-EMBLrel. 16, Last sequence update)				
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)				
DE	Pollen major allergen 2 protein precursor.				
GN	UN2.				
OS	Juniperus ashei (Ozark white cedar).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Coniferopsida; Cupressaceae; Juniperus.				
OX	NCBI_Taxid=13101;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=MALE POLLEN;				
RC	MEDLINE=20403896; PubMed=10944464;				
RA	Yokoyama M., Miyahara M., Shimizu K., Kino K., Tsunoo H.;				
RT	"Purification, Identification and cDNA cloning of Jun a 2, the second				
RT	major allergen of mountain cedar pollen."				
RL	Biochem. Biophys. Res. Commun. 275:195-202 (2000).				
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES (POLYGLACTURONASES).				
CC	EMBL; AJ404653; CAC05582.1; -				
DR	HSSP; P26509; 1BHE.				
DR	InterPro; IPR000743; GH28.				
DR	InterPro; IPR000408; Reg_cnr-condens.				
DR	Pfam; PF00295; Glyco_hydro_28; 1.				
DR	PROSITE; PS00502; POLYGLACTURONASB; UNKNOWN_1.				
DR	PROSITE; PS00626; RCCL_2; UNKNOWN_1.				
KW	Cell wall; Glycosidase; Hydrolase; Signal.				
FT	SIGNAL				
FT	SEQUENCE 507 AA; 55730 MW; 282BOAA5B58FE5A CRC64;				
QY	Query Match	68.4%;	Score 54;	DB 10;	Length 507;
	Best Local Similarity	71.4%;	Pred. No. 0.2;		
	Matches 10;	Conservative 2;	Mismatches 2;	Indels 0;	Gaps 0;
	1 GIDIFASKNFHLQKN 14				

Db 236 GIDIFASKNFHLOK 249

## RESULT 2

Q9M6Y2 ID 09M6Y2 PRELIMINARY; PRT; 503 AA.  
AC 09M6Y2;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Lipoprotein lipase.  
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
NCBI\_TaxID=8022;  
OX NCB1\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ADIPOSE TISSUE;  
RA Lindberg A., Oliverson G.;  
RT "Lipoprotein lipase from rainbow trout differs in several respects from the enzyme in mammals."  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=OVARY;  
RA Prat F., Kwon J.Y., Tyler C.R.;  
RT "Molecular characterization of putative yolk processing enzymes and their expression during oogenesis and embryogenesis in rainbow trout Oncorhynchus mykiss."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ224693; CAB40545.1; -  
DR EMBL; AF358669; AAK69707.1; -  
DR HSSP; P06857; IRP1.  
DR InterPro; IPR000734; Lipase.  
DR InterPro; IPR001024; Lipoxigenase\_LH2.  
DR InterPro; IPR000379; Ser-esterase\_site.  
DR Pfam; PF00151; Lipase; 1.  
DR Pfam; PF01477; PLAT; 1.  
DR PRINTS; PR00821; TAGLIPASE.  
DR SMART; SM00308; LH2; 1.  
DR PROSITE; PS00120; LIPASE\_SER; 1.  
SQ SEQUENCE 503 AA; 57273 MW; FDB3BA08C9D0CA66 CRC64;  
Query Match 57.0%; Score 45; DB 13; Length 503;  
Best Local Similarity 58.3%; Pred. No. 8.5;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 3 DIFASKNFHLOK 14  
DB 429 DIFASKNFHLOK 440  
RESULT 3  
Q9AHX1 ID 09AHX1 PRELIMINARY; PRT; 252 AA.  
AC 09AHX1;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE ATP synthase gamma subunit.  
GN ATPG.  
OS Carsonella ruddii.  
OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
OX NCB1\_TaxID=114186;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=2115546; PubMed=11222582;  
RA Clark M.A., Baumann L., Thao M.L., Moran N.A., Baumann P.;  
RT "Degenerative Minimalism in the genome of a Psyllid Endosymbiont."  
RL EMBL; AF291051; AAK17110.1; -

DR InterPro; IPR00131; ATPase\_gamma.  
DR Pfam; PF00231; ATP\_Synt\_1.  
DR PRINTS; PR00126; ATPASEGAMMA.  
SQ SEQUENCE 252 AA; 29902 MW; 450A3F71C8039A8D CRC64;  
Query Match 54.4%; Score 43; DB 2; Length 252;  
Best Local Similarity 50.0%; Pred. No. 9.4;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLOK 14  
DB 106 GIDIFASKNFHLOK 119

## RESULT 4

Q15807 ID 015807 PRELIMINARY; PRT; 1360 AA.  
AC 015807;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE ATP-binding cassette protein 1 (Fragment).  
GN ABC1.  
OS Cryptosporidium parvum.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
OC Cryptosporidiidae; Cryptosporidium.  
OX NCB1\_TaxID=5807;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KSU-1;  
RA MEDLINE=97376910; PubMed=9233681;  
RX Perkins M.E., Volkman S., Wirth D.F., le Blancq S.M.;  
RT "Characterization of an ATP-binding cassette transporter in Cryptosporidium parvum."  
RL Mcl. Biochem. Parasitol. 87:117-122(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KSU-1;  
RA Zapata F., Riosas Y.A., Wu T.W., le Blancq S.M.;  
RT "ABC proteins in Cryptosporidium parvum."  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KSU-1;  
RA leBlancq S.M., Riosas Y.A., Wu T.W.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U90628; AAC47653.2; -  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR001140; ABCtransporter\_TM.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00664; ABC\_membrane; 1.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR Prodom; PD000006; ABC\_transporter; 3.  
DR SMART; SM00382; AAA\_2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 2.  
DR ATP-binding.  
KW NON TER 1360 1360  
SQ SEQUENCE 1360 AA; 154147 MW; 4ECD0DAD43521994 CRC64;  
Query Match 54.4%; Score 43; DB 5; Length 1360;  
Best Local Similarity 61.5%; Pred. No. 56;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 3 DIFASKNFHLOK 15  
DB 453 DIFASKNFHLOK 465  
RESULT 5  
Q9XYH6 ID 09XYH6 PRELIMINARY; PRT; 1431 AA.  
AC 09XYH6;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, last sequence update)  
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 GN CPABC.  
 OS Cryptosporidium parvum.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 CC Cryptosporidiidae; Cryptosporidium.  
 OX NCBI\_TaxID=5807;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SF6H1;  
 RX MEDLINE=99254110; PubMed=10318953;  
 RA Perkins M.E., Riojas V.A., Wu T.W., Le Blancq S.M.;  
 RT "OpABC, a cryptosporidium parvum ATP-binding cassette protein at the  
 host-parasite boundary in intracellular stages.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 96:5734-5739(1999).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; AF110147; AAD32093.1; -  
 DR HSP; P13569; INBD.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR001140; ABCtransporterTM.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00664; ABC\_membrane; 2.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transport; 2.  
 DR SMART; SM00382; AAA\_2  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 2.  
 DR ATP-binding; Transport.  
 KW SEQUENCE 1431 AA; 162177 MW; 18ABAD9184EB4E CRC64;  
 SQ

Query Match 54.4%; Score 43; DB 5; Length 1431;  
 Best Local Similarity 61.5%; Pred. No. 59;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 IDIFASKNFHL 15  
 DB 453 DILSINFNIQKN 465

RESULT 6  
 ID Q9PQ31 PRELIMINARY; PRT; 156 AA.  
 AC Q9PQ31;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, last annotation update)  
 DE Hypothetical protein U0459.  
 GN U0459.  
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 CC Mycoplasmataceae; Ureaplasma.  
 OX NCBI\_TaxID=134821;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SEROVAR 3;  
 RX MEDLINE=20500219; PubMed=11048724;  
 RA Glaes J.I., Lefkowitz E.J., Glaes J.S., Heiner C.R., Chen E.Y.,  
 RA Casella G.H.;  
 RT "The complete sequence of the mucosal pathogen Ureaplasma  
 urealyticum.";  
 RL Nature 407:757-762(2000).  
 DR EMBL; AE002142; AAF30871.1; -  
 CC Hypothetical protein; Complete proteome.  
 KW SEQUENCE 156 AA; 18839 MW; E50C5FA78412CA7 CRC64;  
 SQ

Query Match 53.2%; Score 42; DB 16; Length 156;  
 Best Local Similarity 72.7%; Pred. No. 8.6;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 IDIFASKNFHL 12  
 DB 66 IKIFKTNFHL 76

RESULT 7  
 ID 045343 PRELIMINARY; PRT; 305 AA.  
 AC 045343;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, last annotation update)  
 DE P10A3.12 protein.  
 GN P10A3.12.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lloyd C.R.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none.  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z92829; CAB07348.1; -  
 DR InterPro; IPR003002; 7TM\_chemol.  
 DR InterPro; IPR00168; 7TM\_nematode.  
 DR Pfam; PF01461; 7tm\_4; 1.  
 KW SEQUENCE 305 AA; 35272 MW; DB66CA7B6DC1F4D CRC64;  
 SQ

Query Match 53.2%; Score 42; DB 5; Length 305;  
 Best Local Similarity 72.7%; Pred. No. 18;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 IDIFASKNFHL 12  
 DB 112 IEIFAPKRFHL 122

RESULT 8  
 ID Q6652 PRELIMINARY; PRT; 630 AA.  
 AC Q6652;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE ORF 50.  
 OS Equine herpesvirus type 2 (strain 86/87) (EHV-2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Gammaherpesvirinae.  
 OX NCBI\_TaxID=82831;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Telford E.A.R.;  
 RX MEDLINE=95302501; PubMed=7783207;  
 RA Telford E.A., Watson M.S., Aird H.C., Perry J., Davison A.J.;  
 RT "The DNA sequence of equine herpesvirus 2.";  
 RL J. Mol. Biol. 249:520-528(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Telford E.A.R.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U20824; AAC13638.1; -  
 DR InterPro; IPR004998; Herpes\_TAF50.  
 DR Pfam; PF03326; Herpes\_TAF50; 1.  
 KW SEQUENCE 630 AA; 70500 MW; 1.99BB2B81A7B060B5 CRC64;  
 SQ

Query Match 51.9%; Score 41; DB 12; Length 630;  
 Best Local Similarity 50.0%; Pred. No. 57;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 IDIFASKNFHL 15

Db 614 LDIISLHNLHUKS 627

## RESULT 9

ID 081407 PRELIMINARY; PRT; 251 AA.

AC 081407; PRELIMINARY; PRT; 251 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE 1-aminocyclopropane-1-carboxylate synthase 4 (Fragment).

GN ACS4.

OS Sinapis arvensis (Charlock).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Eurosids II; Brassicales; Brassicaceae; Sinapis.

OX NCBI\_TaxID=29728;

RN [1]

RP SEQUENCE FROM N.A.

RA Wei Y.D., Hall J.C.;

RT "Characterization of the 1-aminocyclopropane-1-carboxylate synthase gene family of Sinapis arvensis."

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

CC -1- CORACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE TWO ISOZYMES: A CYTOPLASMIC ONE AND A MITOCHONDRIAL ONE (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.

CC EMBL; AF074930; AAC32839.1; -.

DR HSSP; P37821; 1B86.

DR InterPro; IPR001176; ACC synthase.

DR InterPro; IPR004839; Aminotransf1/2.

DR Pfam; PF00155; aminotran 1.2; 1.

DR PRINTS; PR00753; ACCSYNTHASE.

DR PROSITE; PS00105; AA\_TRANSFER\_CLASS\_1; 1.

KW Pyridoxal phosphate. 1

FT NON\_TER 1 251

FT NON\_TER 1 251

SQ SEQUENCE 251 AA; 28063 MW; DCD158F54C8B993A CRC64;

Query Match 50.6%; Score 40; DB 10; Length 251;

Best Local Similarity 63.6%; Pred. No. 33;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 IDIFASKNFHL 12

Db 171 VDFIASKNFHL 181

RESULT 10

ID 08XJS9 PRELIMINARY; PRT; 445 AA.

AC 08XJS9; PRELIMINARY; PRT; 445 AA.

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein CPE1675.

GN CPE1675.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI\_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=13 / TYPE A;

RX PubMed=11792842;

RA Shimizu T., Ohceni K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater."

RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

DR EMBL; AP003191; BAB81381.1; -.

DR InterPro; IPR001861; UPF0004.

DR Pfam; PF00919; UPF0004; 1.

DR TIGRFAMs; TIGR01125; UPF0004; 1.

DR TIGRFAMs; TIGR00089; UPF0004; 1.

DR PROSITE; PS01278; UPF0004; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 445 AA; 50621 MW; 4C7720A60F18A390 CRC64;

## Query Match

Best Local Similarity 50.6%; Score 40; DB 16; Length 445;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 GIDIFASKNFHL 12

Db 207 GIDIFASKNFHL 218

## RESULT 11

ID 098PV1 PRELIMINARY; PRT; 598 AA.

AC 098PV1; PRELIMINARY; PRT; 598 AA.

DT 01-OCT-2001 (TREMBlrel. 18, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DE Sugar ABC transporter permease protein.

RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley B.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
 RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liao X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Plattman G.S., Pan S., Pollard J., Put V., Reese M.G.,  
 RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "the genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003566; AAF50781.1; -  
 DR FLYBase; FBgn0035604; CG10633.  
 DR InterPro; IPR001320; Ion\_glu\_receptor.  
 DR Pfam; PF00060; lig\_chan\_1;  
 SQ SEQUENCE 859 AA; 97964 MW; F37BCD462DC2C610 CRC64;

Query Match 50.6%; Score 40; DB 5; Length 859;  
 Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 DIFASKNFHLQK 15  
 DB 81 DAFARNFOLIKS 93

RESULT 13  
 O9XEV7 PRELIMINARY; PRT; 490 AA.  
 AC O9XEV7;  
 DT 01-NOV-1999 (TEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
 DE Hypothetical 54.7 kDa protein.  
 OS *Oryza sativa* (rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; *Oryza*.  
 OC NCBI\_Taxid=4530;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RA Llaça V., Lou A., Young S., Messing J.;  
 RT "Microscentery in a gene-dense region in maize, sorghum, and rice.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF19222; AAD27681.1; -  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR Pfam; PF00651; BTB\_1.  
 DR PRINTS; PRO1574; TUBBYPROTEIN.  
 DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 490 AA; 54682 MW; 65B8327A713385C7 CRC64;

Query Match 50.0%; Score 39.5; DB 10; Length 490;  
 Best Local Similarity 53.3%; Pred. No. 83;  
 Matches 8; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 GID-IFASKNFHLQ 14  
 DB 53 GIEGVFADKSYHLPK 67

RESULT 14  
 O87205 PRELIMINARY; PRT; 64 AA.  
 AC O87205;  
 DT 01-NOV-1998 (TEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)  
 DE Hypothetical 8.2 kDa protein.  
 OS *Lactococcus lactis*.  
 OC Plasmid pMR01.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Lactococcus.  
 OC NCBI\_Taxid=1358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DPC3147;  
 RX MEDLINE=99000510; PubMed=9767571;  
 RA Dougherty B.A., Hall C., Weidman J.F., Richardson D.R., Venter J.C.,  
 RA Ross R.P.;  
 RT "Sequence and analysis of the 60 kb conjugative, bacteriocin-producing  
 RT plasmid pMR01 from *Lactococcus lactis* DPC3147.";  
 RL Mol. Microbiol. 29:1029-1038(1998).  
 DR EMBL; AE001272; AAC56026.1; -  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 64 AA; 8187 MW; 23713F5794FBE304 CRC64;

Query Match 49.4%; Score 39; DB 2; Length 64;  
 Best Local Similarity 61.5%; Pred. No. 12;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLQ 13  
 DB 44 GILIFLKNPHTQ 56

RESULT 15  
 O11321 PRELIMINARY; PRT; 122 AA.  
 AC O11321;  
 DT 01-JUL-1997 (TEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE B-N'.2 protein (Fragment).  
 GN B-N'.2.  
 OS Molluscum contagiosum virus subtype 1 (MCV1).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Molluscipoxvirus.  
 OC NCBI\_Taxid=10280;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RA Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,  
 RA Lopez-Esteban J.L., Esteban M., Martin-Gallardo A.,  
 RT "A Random DNA Sequencing, Computer-Based Approach for the Generation  
 RT of a Gene Map of Molluscum Contagiosum Virus.";

RL Virus Genes 0:0-0(1997).  
DR EMBL: U86902; AAB57951.1; -  
FT NON TER 122 122  
SQ SEQUENCE 122 AA; 14361 MW; 2284E1A5B77492F CRC64;

Query Match 49.4%; Score 39; DB 12; Length 122;  
Best Local Similarity 60.0%; Pred. No. 23;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 DIFASKNEFL 12  
|:|:|:|:|:  
DB 54 DVFANVNFHV 63

Search completed: April 20, 2003, 13:13:16  
Job time : 14.6711 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds

(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524d-121

Perfect score: 79

Sequence: 1 ASKNFHLQKNTIGTG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73:\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*  
5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	514	2	S48730
2	79	100.0	514	2	JC2498
3	54	68.4	514	2	JC7100
4	51	64.6	507	2	JC7366
5	45	57.0	419	2	T38199
6	45	57.0	513	2	T11686
7	43	54.4	97	1	R6RT37
8	43	54.4	97	1	S47646
9	43	54.4	275	2	T28738
10	42	53.2	421	2	B83615
11	41	51.9	491	2	D97037
12	41	51.9	530	1	A27481
13	40	50.6	386	2	E71693
14	40	50.6	549	1	S48474
15	40	50.6	767	1	JU0474
16	40	50.6	778	1	ALBYG
17	39	49.4	104	2	S07705
18	39	49.4	104	2	G24402
19	39	49.4	109	2	F81669
20	39	49.4	115	2	C24891
21	39	49.4	116	2	D24891
22	39	49.4	118	2	I71934
23	39	49.4	121	2	A29080
24	39	49.4	133	2	F24402
25	39	49.4	342	2	A95165
26	39	49.4	346	2	H98030
27	39	49.4	352	2	G81921
28	39	49.4	391	2	T51609
29	39	49.4	723	2	H85092

## ALIGNMENTS

30 39 49.4 874 2 A10281  
31 38.5 48.7 819 2 T30050  
32 38 48.1 60 2 A30606  
33 38 48.1 173 2 JN0515  
34 38 48.1 174 2 S71554  
35 38 48.1 181 2 AF1493  
36 38 48.1 250 2 AB1766  
37 38 48.1 250 2 AB1766  
38 38 48.1 252 2 C97426  
39 38 48.1 262 2 AC2644  
40 38 48.1 324 2 AG2956  
41 38 48.1 344 2 G98326  
42 38 48.1 341 1 DEECTH  
43 38 48.1 341 2 F91190  
44 38 48.1 341 2 G86037  
45 38 48.1 343 2 B82405

probable insectici  
hypothetical prote  
T-cell receptor al  
FliV protein - Sal  
pathogenesis-relat  
hypothetical prote  
regulatory protein  
regulatory protein  
flagellar basal-bo  
flagellar basal-bo  
hypothetical prote  
c4-dicarboxylate-b  
L-threonine 3-dehy  
threonine dehydrog  
threonine dehydrog  
threonine 3-dehydr

RESULT 1  
S48730  
Cry j II protein - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999  
C/Accession: S48730  
R/Namba, M.; Kurose, M.; Torioka, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.;  
FBS Lett. 353, 124-128, 1994  
A/Title: Molecular cloning of the second major allergen, Cry j II, from Japanese ceda  
A/Reference number: S48730; MUID:95010777; PMID:7926035  
A/Accession: S48730  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-514 <NAM>  
A/Cross-References: GB:D37765; NID:9577695; PIDN:BA07021.1; PID:dl007598; PID:9577696

Query Match 100.0%; Score 79; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASKNFHLQKNTIGTG 15  
Db 240 ASKNFHLQKNTIGTG 254

RESULT 2  
JC2498  
second major allergen Cry j II precursor - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000  
C/Accession: JC2498; PC2346; A60147  
R/Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994  
A/Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese  
A/Reference number: JC2498; MUID:94271186; PMID:8002972  
A/Accession: JC2498  
A/Molecule type: mRNA  
A/Residues: 1-514 <KOM>  
A/Cross-References: DDBJ:D29772; NID:9506857; PIDN:BA06172.1; PID:9506858

A/Accession: PC2346  
A/Molecule type: protein  
A/Residues: 52-61 <KO2>  
R/Sakaguchi, M.; Inouye, S.; Tanhai, M.; Ando, S.; Usui, M.; Matuhasi, T.  
Allergy 45, 309-312, 1990  
A/Title: Identification of the second major allergen of Japanese cedar pollen.  
A/Reference number: A60147; MUID:90342988; PMID:2382797  
A/Accession: A60147  
A/Molecule type: protein  
A/Residues: 55-64 <SAK>  
C/Keywords: glycoprotein; pollen  
F1-54/Domain: signal sequence #status predicted <SIG>  
F155-460/Product: second major allergen Cry j #status predicted <MAT>

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 79; DB 2; Length 514;

Best Local Similarity 100.0%; Pred. No. 1.3e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

### RESULT 3

UC7100 Polygalacturonase Cha o 2 - Japanese cypress

C/Species: Chamaecyparis obtusa (Japanese cypress)

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

C/Accession: J07100; PC7026

R/Mori, T.; Yokoyama, M.; Komiyama, N.; Okano, M.; Kino, K.

Biochem. Biophys. Res. Commun. 263, 166-171, 1999

A/Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A/Reference number: J07100; MUID:99417540; PMID:10486272

A/Accession: J07100

A/Molecule type: mRNA

A/Residues: 1-514 <MOR>

A/Accession: PC7026

A/Molecule type: protein

A/Residues: 51-62 <MO2>

Query Match 68.4%; Score 54; DB 2; Length 514;

Best Local Similarity 66.7%; Pred. No. 0.061;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 240 ASKNFHLQKNTIGTG 254

### RESULT 4

UC7366 Jun a 2 protein - mountain cedar

C/Species: Juniperus ashei (mountain cedar)

C/Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C/Accession: J07366; PC7093

R/Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.

Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A/Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A/Reference number: J07366

A/Accession: J07366

A/Molecule type: mRNA

A/Residues: 1-507 <YOK>

A/Cross-references: GB:AJ04653

A/Accession: PC7093

A/Molecule type: protein

A/Residues: 55-63 <Y02>

C/Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

C/Keywords: glycoprotein; pollen

Query Match 64.8%; Score 51; DB 2; Length 507;

Best Local Similarity 66.7%; Pred. No. 0.22;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASKNFHLQKNTIGTG 15  
DB 241 ASKNFHLQKNTIGTG 255

### RESULT 5

T8189 conserved hypothetical protein SPAC22F8.09 - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C/Accession: T8189

R/Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Pohl, T.

submitted to the EMBL Data Library, August 1999

A/Reference number: Z21777

A/Accession: T8189

A/Status: preliminary; translated from GB/EMBL/DDB

A/Molecule type: DNA

A/Residues: 1-419 <LYN>

A/Cross-references: EMBL:AL098311; PIR:GAB52719.1; GSPDB:GN00066; SPDB:SPAC22F8.09

A/Experimental source: strain 972h-, cosmid c22F8

C/Genetics: SPDB:SPAC22F8.09

A/Map position: 1

A/Introns: 46/2

Query Match 57.0%; Score 45; DB 2; Length 419;

Best Local Similarity 80.0%; Pred. No. 2.3;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 KNFHLQKNTI 12  
DB 410 KDFHLQKNTI 419

### RESULT 6

T11686 aspartic proteinase (EC 3.4.23.-) - cowpea

C/Species: Vigna unguiculata (cowpea)

C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C/Accession: T11686

R/d'Arcy-Jamara, A.; Zully-Fodil, Y.; Pham Thi, A.T.; Ferrarri-Ilhou, R.

submitted to the EMBL Data Library, June 1996

A/Description: Molecular cloning and nucleotide sequence of a cDNA encoding an aspartic

A/Reference number: Z17314

A/Accession: T11686

A/Status: preliminary; translated from GB/EMBL/DDB

A/Molecule type: mRNA

A/Residues: 1-513 <DAR>

A/Cross-references: EMBL:U61396; NID:G1420935; PID:G1420936

A/Experimental source: strain 1183

C/Superfamily: oryzasin; saposin repeat homology

C/Keywords: aspartic proteinase; hydrolase

Query Match 57.0%; Score 45; DB 2; Length 513;

Best Local Similarity 61.5%; Pred. No. 2.9;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 KNFHLQKNTIGTG 15  
DB 62 KDFHLQKNTIGTG 74

### RESULT 7

R6R737 ribosomal protein L37, cytosolic [validated] - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 25-Feb-1995 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000

C/Accession: J00478; P00455; S21496; A02779; S11421

R/Chan, Y.L.; Paz, V.; Olivera, J.; Wool, I.G.

Biochem. Biophys. Res. Commun. 192, 590-596, 1993

A/Title: The primary structure of L37 - a rat ribosomal protein with a zinc finger-1

A/Reference number: J00478; MUID:93249430; PMID:8484768

A/Accession: J00478

A/Molecule type: mRNA

A/Residues: 1-97 <CHA>

A/Accession: P00455

A/Molecule type: protein

A/Residues: 2-40 <CH1>

A/Note: The protein is designated as ribosomal protein L37

R/Chan, Y.L.; Paz, V.; Olivera, J.; Wool, I.G.

submitted to the EMBL Data Library, May 1992

A/Description: The primary structure of rat ribosomal protein S9.

A/Reference number: S21496

A/Accession: S21496

A/Molecule type: mRNA  
 A/Residues: 1-95 <CHW>  
 C/Cross-references: EMBL:X66369  
 R/Lin, A.; McNally, J.; Wool, I.G.  
 J. Biol. Chem. 258, 10664-10671, 1983  
 A/Title: The primary structure of rat liver ribosomal protein L37. Homology with yeast  
 A/Reference number: A02779; MUID:83291000; PMID:6350292  
 A/Accession: A02779  
 A/Molecule type: protein  
 A/Residues: 'N', 3-18, 'G', 20-21, 'KSKGKALOKGNSFPEKL', 22-25, 26-30, 'SKKNTLHAKRTTGM', 40-4  
 A/Note: The protein is designated as ribosomal protein L37  
 R/Wittmann-Liebhold, B.; Geisler, A.W.; Lin, A.; Wool, I.G.  
 J. Supramol. Struct. 12, 425-433, 1979  
 A/Title: Sequence of the amino-terminal region of rat liver ribosomal proteins S4, S6, S  
 A/Reference number: S11413; MUID:80252792; PMID:398910  
 A/Accession: S11421  
 A/Molecule type: protein  
 A/Residues: 'X', 3-18, 'G', 20-21, 'KKKGKA', 29-31 <MT>  
 A/Note: The protein is designated as ribosomal protein L37  
 C/Superfamily: rat ribosomal protein L37  
 C/Keywords: cytosol; protein biosynthesis; ribosome; zinc finger  
 F/2-97/Product: ribosomal protein L37 #status experimental <RPL>  
 F/19-37/Region: zinc finger CCCC motif

Query Match 54.4%; Score 43; DB 1; Length 97;  
 Best Local Similarity 66.7%; Pred. No. 1;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 SKNPHLOKNTIG 13  
 DB 24 SKAHLKSTCG 35

RESULT 8  
 S47646  
 ribosomal protein L37, cytosolic - human  
 C/Species: Homo sapiens (man)  
 C/Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 13-Aug-1999  
 C/Accession: S47646; S68344  
 R/Barnard, G.F.; Stalinas, R.J.; Puder, M.; Steele Jr., G.D.; Chen, L.B.  
 Biochim. Biophys. Acta 1218, 425-428, 1994  
 A/Title: Human ribosomal protein L37 has motifs predicting serine/threonine phosphorylat  
 A/Reference number: S47646; MUID:94325352; PMID:7545944  
 A/Accession: S47646  
 A/Molecule type: mRNA  
 A/Residues: 1-97 <BAR>  
 A/Cross-references: EMBL:L11567; NID:G292440; PIDN:AA62148.1; PID:G292441  
 R/Su, S.; Bird, R.C.  
 Eur. J. Biochem. 232, 789-797, 1995  
 A/Title: Cell cycle, differentiation and tissue-independent expression of ribosomal prot  
 A/Reference number: S68344; MUID:96028120; PMID:7588717  
 A/Accession: S68344  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-97 <SUS>  
 A/Cross-references: GB:S79979; NID:G1839333; PIDN:AA647039.1; PID:G1839334  
 C/Genetics: GDB:RPL37  
 A/Map position: 14pter-14qter  
 C/Superfamily: rat ribosomal protein L37  
 C/Keyword: protein biosynthesis; ribosome; zinc finger

Query Match 54.4%; Score 43; DB 2; Length 97;  
 Best Local Similarity 66.7%; Pred. No. 1;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 SKNPHLOKNTIG 13  
 DB 24 SKAHLKSTCG 35

RESULT 9

T28738  
 hypothetical protein F48G7.4 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C/Accession: T28738  
 R/Glarke, K.; Mohlmann, P.; Harrison, M.  
 submitted to the EMBL Data Library, January 1998  
 A/Description: The sequence of C. elegans cosmid F48G7.  
 A/Reference number: Z20517  
 A/Accession: T28738  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-275 <CIA>  
 A/Cross-references: EMBL:AF039044; PIDN:AA647944.1; GSPDB:GN00023; CESP:F48G7.4  
 A/Experimental source: strain Bristol NZ; clone F48G7  
 C/Genetics:  
 A/Gene: CESP:F48G7.4  
 A/Map position: 5  
 A/Introns: 189/3; 232/2

Query Match 54.4%; Score 43; DB 2; Length 275;  
 Best Local Similarity 64.3%; Pred. No. 3.4;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 SKNPHLOKNTIG 15  
 DB 230 SKRFYQKQIGTG 243

RESULT 10  
 B83615  
 probable porin PA0240 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Jun-2001  
 C/Accession: B83615  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 ; Loy, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat  
 A/Reference number: B83615  
 A/Accession: B83615  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-421 <STO>  
 A/Cross-references: GB:AE004462; GB:AE004091; NID:G9946077; PIDN:AA603629.1; GSPDB:GN  
 A/Experimental source: strain PA01  
 C/Genetics:  
 A/Gene: PA0240  
 C/Superfamily: Pseudomonas aeruginosa outer membrane porin PA0958

Query Match 53.2%; Score 42; DB 2; Length 421;  
 Best Local Similarity 50.0%; Pred. No. 8.5;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 SKNPHLOKNTIG 13  
 DB 43 ARNPHLRNFGV 54

RESULT 11  
 D97037  
 hypothetical protein CAC113 [imported] - Clostridium acetobutylicum  
 C/Species: Clostridium acetobutylicum  
 C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C/Accession: D97037  
 R/Nolling, U.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Le  
 ; Daly, M.J.; Benoit, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C  
 A/Reference number: A96900; MUID:21359325; PMID:21359325  
 A/Accession: D97037  
 A/Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-491 <KOR>  
A:Cross-references: GB:AE001437; PIDN:AAK79087.1; PID:g15024032; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC113

Query Match 51.9%; Score 41; DB 2; Length 491;  
Best Local Similarity 57.1%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 SKNFHLOKNTIGTG 15  
DB 326 SWNYHLRNTIGTG 339

## RESULT 12

A27491  
lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - yeast (Saccharomyces

N:Alternate names: protein YHR007C  
C:Species: Saccharomyces cerevisiae  
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 23-Mar-2001  
C:Accession: A27491; S46804; B31569; A25563

R:Kahl, V.F.; Woods, C.W.; Turf, T.G.; Dey, C.R.; Sutter, T.R.; Loper, J.C.  
DNA 6, 529-537, 1987

A:Title: Primary structure of the P450 lanosterol demethylase gene from Saccharomyces ce  
A:Reference number: A27491; MUID:88111027; PMID:3322742  
A:Accession: A27491

A:Molecule type: DNA  
A:Residues: 1-530 <KAL1>  
A:Cross-references: EMBL:M18109; NID:g170945; PIDN:AAA34379.1; PID:g170946

R:Favelli, T.  
Submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of S. cerevisiae cosmid 9780.

A:Reference number: S46797  
A:Accession: S46804

A:Molecule type: DNA  
A:Residues: 1-530 <PAV>

A:Cross-references: EMBL:U10555; NID:g500813; PIDN:AA68433.1; PID:g500824; GSPDB:GN0000  
R:Ikshida, N.; Aoyama, Y.; Hatanaka, R.; Oyama, Y.; Inajo, S.; Ishiguro, M.; Oshima, T.;  
Biochem. Biophys. Res. Commun. 155, 317-323, 1988

A:Title: A single amino acid substitution converts cytochrome P450-14DM to an inactive H  
A:Reference number: A31569; MUID:88326319; PMID:3046615  
A:Accession: B31569

A:Molecule type: DNA  
A:Residues: 1-432, 'N', 434-530 <ISH>  
A:Cross-references: EMBL:M21463; NID:g171353; PIDN:AAA34546.1; PID:g171354

R:Kahl, V.F.; Loper, J.C.; Dey, C.R.; Woods, C.W.; Sutter, T.R.  
Gene 45, 237-245, 1986  
A:Title: Isolation of a cytochrome P-450 structural gene from Saccharomyces cerevisiae.

A:Reference number: A25563; MUID:87106820; PMID:3542713  
A:Accession: A25563

A:Molecule type: DNA  
A:Residues: 444-530 <KAL2>  
A:Cross-references: EMBL:M15653

A:Genetics: SGD:ERG11; MIPS:YHR007C  
A:Cross-references: SGD:S0001049; MIPS:YHR007C

A:Map position: 8R  
C:Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F:311492/Domain: cytochrome P450 homology <P45>  
F:470/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 51.9%; Score 41; DB 1; Length 530;  
Best Local Similarity 51.5%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 SKNFHLOKNTIGTG 14  
DB 181 SKNFRLNRTIGTG 193

## RESULT 13

E71693  
hypothetical protein RP364 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii  
C>Date: 21-Nov-1998 #sequence revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: E71693

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark,  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: E71693

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-386 <AND>  
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CA14823.1; PID:g3866  
A:Experimental source: strain Madrid E  
C:Genetics:

A:Gene: RP364

Query Match 50.6%; Score 40; DB 2; Length 386;  
Best Local Similarity 54.5%; Pred. No. 18;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 FHLOKNTIGTG 15  
DB 214 FHLOKNTIGTG 224

## RESULT 14

S48474  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) SGAL - yeast (Saccharomyces cerevisiae)

N:Alternate names: glucanase; protein YIL099W  
C:Species: Saccharomyces cerevisiae  
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 12-Nov-1999

C:Accession: S48474; C26877; S27284  
R:Bowman, S.; Churche, C.  
Submitted to the EMBL Data Library, September 1994

A:Reference number: S48455  
A:Accession: S48474

A:Molecule type: DNA  
A:Residues: 1-549 <BOW>

A:Cross-references: GB:247047; EMBL:Z38125; NID:g603997; PID:g763247; GSPDB:GN00009;  
R:Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987

A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.  
A:Reference number: A91831; MUID:87194600; PMID:3106330  
A:Accession: C26877

A:Molecule type: DNA  
A:Residues: 1-503,507, 'W', 513-514, 'TG', 516 <YAM>

A:Cross-references: EMBL:M16166; NID:g172592; PIDN:AAA35042.1; PID:g172593  
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988

A:Title: Similar short elements in the 5' regions of the STAL and SGA genes from Sacc  
A:Reference number: S27281; MUID:89031230; PMID:3141213  
A:Accession: S27284

A:Molecule type: DNA  
A:Residues: 1-183, 'H', 185-190 <PAR>

A:Cross-references: EMBL:X13858; NID:g4461; PIDN:CAA32071.1; PID:g4463  
C:Genetics: SGD:SGAL; MIPS:YIL099W  
A:Cross-references: SGD:S0001361; MIPS:YIL099W

A:Map position: 9L  
C:Function:

A:Description: hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase SGAL; glucan 1,4-alpha-glucosidase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; yeast vacuole  
F:77-536/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 50.6%; Score 40; DB 1; Length 549;  
Best Local Similarity 57.1%; Pred. No. 27;  
Matches 8; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 4 NFHLOK-NTIGTG 15

Db 159 SFHLQSRNNTLGAG 172

## RESULT 15

JU0474

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) STA2 - yeast (Saccharomyces diastaticus)

N/Alternate names: glucoamylase II

C/Species: Saccharomyces diastaticus

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C/Accession: JU0474; S35895

R/Lambrechts, M.G.; Pretorius, I.S.; Sollitt, P.; Marmur, J.

Gene 100, 95-103, 1991

A/Title: Primary structure and regulation of a glucoamylase-encoding gene (STA2) in Sacc

A/Reference number: JU0474; MUID:91276266; PMID:2055484

A/Accession: JU0474

A/Molecule type: DNA

A/Residues: 1-767 &lt;LAM&gt;

A/Cross-references: GB:M60650; NID:g172733; PIDN:AAA35107.1; PID:g172734

A/Note: the authors translated the codon CCG for residue 337 as Ala, CCG for residue 364

R/Kim, K.; Bajszar, G.; Lee, S.Y.; Knudsen, F.; Mattoon, J.R.

submitted to the EMBL Data Library, May 1992

A/Description: Cloning of a new allelic variant of a Saccharomyces diastaticus glucoamyl

A/Reference number: S35895

A/Accession: S35895

A/Molecule type: DNA

A/Residues: 1-163, 'A', 164-622, 'D', 624-767 &lt;KIM&gt;

A/Cross-references: EMBL:M90490; NID:g172735; PIDN:AAA20560.1; PID:g172736

C/Genetics:

A/Gene: STA2; DEX1

C/Superfamily: yeast glucan 1,4-alpha-glucosidase STA1; glucan 1,4-alpha-glucosidase hom

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

F;334-767/Domain: glucan 1,4-alpha-glucosidase homology &lt;GAG&gt;

Query Match 50.6%; Score 40; DB 1; Length 767;

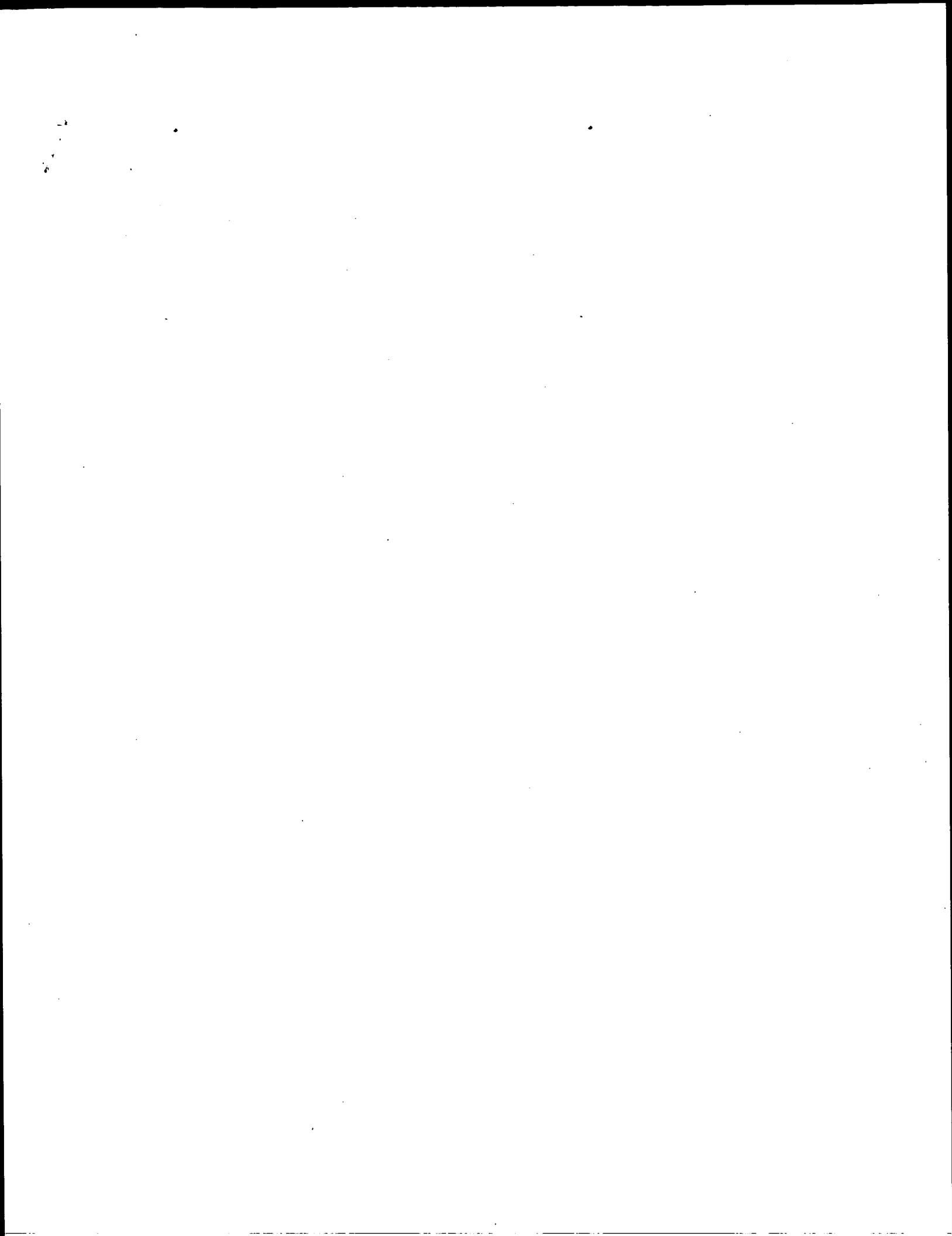
Best Local Similarity 57.1%; Pred. No. 40;

Matches 8; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 4 NFHLQK--NTIGTG 15

Db 416 SFHLQRTNTLTGAG 429

Search completed: April 20, 2003, 13:15:55  
 Job time: 7.07895 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-121

Perfect score: 79

Sequence: 1 ASKNFHLQKNTIGTG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	79	100.0	514	1	MPA2_CRYUA
2	45	57.0	419	1	YIY9_SCHPO
3	43	54.4	92	1	RU37_BOVIN
4	43	54.4	96	1	RU37_HUMAN
5	43	54.4	96	1	RL37_ICTPU
6	41	51.9	530	1	CP51_YEAST
7	40	50.6	385	1	Y364_RICPR
8	40	50.6	549	1	AMYG_YEAST
9	40	50.6	767	1	AMYH_YEAST
10	40	50.6	768	1	AMYH_SACDI
11	39	49.4	234	1	A29B_DROME
12	39	49.4	342	1	QUEA_STREY
13	39	49.4	342	1	QUEA_STREY
14	39	49.4	437	1	PRPG_LACLU
15	38	48.1	173	1	FLIV_SALMU
16	38	48.1	262	1	FLUG_AGRIS
17	38	48.1	341	1	TDH_ECOLI
18	38	48.1	401	1	FLIB_SALCH
19	38	48.1	401	1	FLIB_SALCH
20	37	46.8	64	1	RL35_HELPY
21	37	46.8	262	1	FLUG_RHIME
22	37	46.8	371	1	LE3B_ASPEG
23	37	46.8	380	1	S18C_HUMAN
24	37	46.8	571	1	EXOI_SCHPO
25	37	46.8	657	1	IVRB_MYCPN
26	37	46.8	1034	1	GCSP_YEAST
27	37	46.8	1305	1	RKPL_YEAST
28	36.5	46.2	416	1	EXNI_XENIA
29	36.5	46.2	608	1	DEXT_PENNI
30	36	45.6	85	1	UI86_DROME
31	36	45.6	86	1	UI86_HUMAN
32	36	45.6	86	1	UI86_MOUSE
33	36	45.6	132	1	TVA3_MOUSE

34	36	45.6	200	1	VATE_BORBU	051123 borrelia bu
35	36	45.6	380	1	YI03_MYCPN	P75530 mycoplasma
36	36	45.6	280	1	CYBP_DROME	P75530 mycoplasma
37	36	45.6	341	1	QUEA_CLOAB	Q9V393 drosophila
38	36	45.6	341	1	QUEA_STRAM	Q9V393 drosophila
39	36	45.6	342	1	QUEA_CAMJE	Q9V393 campylobact
40	36	45.6	342	1	QUEA_LISMO	Q92b13 listeria in
41	36	45.6	342	1	QUEA_LISMO	Q92b13 listeria in
42	36	45.6	344	1	TDH_RHIME	052998 rhizobium m
43	36	45.6	345	1	QUEA_LACLA	Q9C146 lactococcus
44	36	45.6	346	1	QUEA_MEIMA	Q9JW05 neisseria m
45	36	45.6	346	1	QUEA_MEIMA	Q9JW05 neisseria m

## ALIGNMENTS

```

RESULT 1
MPA2_CRYUA
ID MPA2_CRYUA STANDARD: PRT; 514 AA.
AC P43212;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Possible polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)
DE (Major pollen allergen Cry j 2) (Cry j II).
OS Cryptomeria japonica (Japanese cedar).
OC Bukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
OC Spermatophyta, Coniferopsida, Coniferales, Cupressaceae, Cryptomeria.
OX NCBI_TaxID=3369;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Pollen;
RX MEDLINE=95010777; PubMed=7926035;
RA Nanda M., Kurose M., Totigoe K., Hino K., Taniguchi Y., Fukuda S.,
RA Usui M., Kurimoto M.;
RT "Molecular cloning of the second major allergen, Cry j II, from
RT Japanese cedar pollen."
RL FEBS Lett. 353:124-128(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen;
RX MEDLINE=94271186; PubMed=8002972;
RA Komiyama N., Some T., Shimizu K., Morikubo K., Kino K.;
RT "cDNA cloning and expression of Cry j II the second major allergen of
RT Japanese cedar pollen."
RL Biochem. Biophys. Res. Commun. 201:1021-1028(1994).
RN [3]
RP SEQUENCE OF 55-64.
RX MEDLINE=90342988; PubMed=2382797;
RA Sakauchi M., Inoue S., Tanai M., Ando S., Usui M., Matubasi T.;
RT "Identification of the second major allergen of Japanese cedar
RT pollen."
RL Allergy 45:309-312(1990).
-1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-
galactosiduronic linkages in pectate and other galacturonans.
-1- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).
-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
(POLYGALACTURONASES).
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: D37765; BAA07021.1; -
EMBL: D29772; BAA06172.1; -
HSSP: P25509; 1BHE.
InterPro: IPR000743; GH28.
Pfam: PF00295; Glyco_hydro_28; 1.

```

DR PROSITE: PS00502; POLYGALACTURONASE; 1.  
 KM Hydrolase; Glycosidase; Cell wall; Signal; Zymogen; Fruit ripening;  
 KW Amyloplast; Glycoprotein; Allergen.  
 FT SIGNAL 1 22  
 FT PROPEP 23 45  
 FT CHAIN 46 433  
 FT PROPEP 434 514  
 FT ACT SITE 278 278  
 FT CARBOHYD 460 460  
 FT CARBOHYD 472 472  
 FT CARBOHYD 5 5  
 FT CONFLICT 12 12  
 FT CONFLICT 34 35  
 FT CONFLICT 37 37  
 FT CONFLICT 88 88  
 FT CONFLICT 98 98  
 FT CONFLICT 451 451  
 FT CONFLICT 454 454  
 FT CONFLICT 504 504  
 FT CONFLICT 507 507  
 SQ SEQUENCE 514 AA; 56645 MW; 624611C3FAD6302 CRC64;  
 Query Match 100.0%; Score 79; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 ASKNFHLQKNTIGTG 15  
 Db 240 ASKNFHLQKNTIGTG 254  
 RESULT 2  
 ID Y1Y9 SCHPO STANDARD; PRT; 419 AA.  
 AC Q9U014;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C22F8.09 in chromosome 1.  
 GN SPAC22F8.09.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NX NCBI\_TaxID=4896;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RC MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Pest N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowicz E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymopiez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Moutier S.,  
 RA Galibert F., Ayes S.J., Xiang Z., Hunt C., Moore K., Huest S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Bente G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.U., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Polishkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- SIMILARITY: BELONGS TO THE GITSR2 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL109831; CAB52719.1; -  
 DR Hypothetical protein.  
 SQ SEQUENCE 419 AA; 48129 MW; 9C2A72BE39E61DNA CRC64;  
 Query Match 57.0%; Score 45; DB 1; Length 419;  
 Best Local Similarity 80.0%; Pred. No. 0.77;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Oy 3 KNFHLQKNTI 12  
 Db 410 KNFHLQKNTI 419  
 RESULT 3  
 ID RL37 BOVIN STANDARD; PRT; 92 AA.  
 AC P79244;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE 60S ribosomal protein L37 (Fragment).  
 GN RPL37.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 NX NCBI\_TaxID=9913;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96028120; PubMed=7588717;  
 RA Su S., Bird R.C.;  
 RT "Cell cycle, differentiation and tissue-independent expression of  
 RT ribosomal protein L37.";  
 RL Eur. J. Biochem. 232:789-797(1995).  
 CC -1- SIMILARITY: BELONGS TO THE L37E FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S79980; AAD14319.1; -  
 DR InterPro: IPR001569; Ribosomal L37E.  
 DR Pfam: PF01907; Ribosomal\_L37E; 1.  
 DR ProDom: PD005132; Ribosomal\_L37E; 1.  
 DR PROSITE; PS01077; RIBOSOMAL\_L37E; PARTIAL.  
 KW Ribosomal protein.  
 FT INIT MET 0  
 FT NON TER 92  
 SQ SEQUENCE 92 AA; 10576 MW; 881F2879AA8E0CF8 CRC64;  
 Query Match 54.4%; Score 43; DB 1; Length 92;  
 Best Local Similarity 66.7%; Pred. No. 0.34;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 SKAYHLQKSTG 13  
 Db 23 SKAYHLQKSTG 34



RESULT 4  
ID RL37\_HUMAN STANDARD; PRT; 96 AA.  
AC P02403; O99883;  
DT 21-JUN-1986 (Rel. 01, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 60S ribosomal protein L37 (G1.16).  
GN RPL37.  
OS Homo sapiens (Human), and  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606, 10116;  
RN [1]  
RP SPECIES=Human; TISSUE=Lymphoma;  
RA Kato S.;  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SPECIES=Human; TISSUE=Colon;  
RX MEDLINE=94325352; PubMed=7545944;  
RA Barnard G.F., Stanlunas R.J., Puder M., Steele G.D. Jr., Chen L.B.;  
RT "Human ribosomal protein L37 has motifs predicting serine/threonine  
phosphorylation and a zinc-finger domain.";  
RL Biochim. Biophys. Acta 1218:425-428(1994).  
RN [3]  
RP SPECIES=Human, and Rat;  
RX MEDLINE=96028120; PubMed=7588717;  
RA Su S., Bird R.C.;  
RT "Cell cycle, differentiation and tissue-independent expression of  
ribosomal protein L37.";  
RL Eur. J. Biochem. 232:789-797(1995).  
RN [4]  
RP SPECIES=Human;  
RX MEDLINE=98248690; PubMed=9582194;  
RA Kennedy N., Kawaguchi T., Rozen S., Davis E., Goodman N.,  
RA Hudson T.J., Tanaka T., Page D.C.;  
RT "A map of 75 human ribosomal protein genes: sequencing and comparative  
analysis of 73 genes.";  
RL Genome Res. 12:379-390(2002).  
RN [5]  
RP SPECIES=Human;  
RX MEDLINE=98248690; PubMed=9582194;  
RA Kennedy N., Kawaguchi T., Rozen S., Davis E., Goodman N.,  
RA Hudson T.J., Tanaka T., Page D.C.;  
RT "A map of 75 human ribosomal protein genes.";  
RL Genome Res. 8:509-523(1998).  
RN [6]  
RP SPECIES=Human, and Rat;  
RX MEDLINE=93249430; PubMed=8484768;  
RA Chan Y.-L., Paz V., Olivera J., Wool I.G.;  
RT "The primary structure of L37 -- a rat ribosomal protein with a zinc  
finger-like motif.";  
RL Biochem. Biophys. Res. Commun. 192:590-596(1993).  
RN [7]  
RP SPECIES=Human;  
RX MEDLINE=83291000; PubMed=6350292;  
RA Lin A., McNally J., Wool I.G.;  
RT "The primary structure of rat liver ribosomal protein L37. Homology  
with yeast and bacterial ribosomal proteins.";  
RL J. Biol. Chem. 258:10664-10671(1983).  
CC -1- SIMILARITY: BELONGS TO THE L37E FAMILY OF RIBOSOMAL PROTEINS.  
CC -1- CAUTION: REF. 7 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; D23661; BA04888.1; -;  
DR EMBL; L11567; AAB62148.1; -;  
DR EMBL; S79979; AAB47039.1; -;  
DR EMBL; S79981; -; NOT ANNOTATED\_CDS.  
DR EMBL; AB061834; BAB79472.1; -;  
DR EMBL; AB007184; BAA25843.1; -;  
DR EMBL; AB007183; BAA25842.1; -;  
DR EMBL; X66369; CAA47012.1; -;  
DR PIR; A02779; R6RT37.  
DR PIR; JN0478; JN0478.  
DR PIR; S21496; S21496.  
DR Genew; HGNC:10347; RPL37.  
DR MIM; 604181; -;  
DR InterPro; IPR001569; Ribosomal\_L37E.  
DR Pfam; PF01907; Ribosomal\_L37e; 1.  
DR ProDom; PD005132; Ribosomal\_L37E; 1.  
DR PROSITE; PS01077; RIBOSOMAL\_L37E; 1.  
DR KMW Ribosomal protein.  
FT INIT MET 0  
FT CONFICT 0  
FT SEQUENCE 96 AA; 10947 MW; 959C318791ADDE39 CRC64;  
SQ  
Query Match 54.4%; Score 43; DB 1; Length 96;  
Best Local Similarity 66.7%; Pred. No. 0.35;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DR 2 SKNFHLOKNTIG 13  
DB 23 SKAYHLOKSTCG 34  
RESULT 5  
ID RL37 ICTPU STANDARD; PRT; 96 AA.  
AC Q90YTL;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 60S ribosomal protein L37.  
GN RPL37.  
OS Ictalurus punctatus (Channel catfish).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
OC Ictaluridae; Ictalurus.  
OX NCBI\_Taxid=7998;  
RN [1]  
RP SPECIES=Human, and Rat;  
RA Patterson A.P., Karai A., Liu Z.J.;  
RT "Translational machinery of channel catfish: II. Complementary DNA and  
expression of the complete set of 47 60S ribosomal proteins.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE L37E FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AF401593; AAK95165.1; -;  
DR InterPro; IPR001569; Ribosomal\_L37E.  
DR Pfam; PF01907; Ribosomal\_L37e; 1.  
DR ProDom; PD005132; Ribosomal\_L37E; 1.

DR PROSITE; PS01077; RIBOSOMAL\_L37E; 1.  
 KM Ribosomal protein.  
 FT INT MET 0 BY SIMILARITY.  
 SQ SEQUENCE 96 AA; 10911 MW; 94A618F26055E983 CRC64;  
 Query Match 54.4%; Score 43; DB 1; Length 96;  
 Best Local Similarity 66.7%; Pred. No. 0.35;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 SKNPHLOXNTIG 13  
 23 SKNPHLOXNTIG 34

RESULT 6  
 ID\_CPS1\_YEAST STANDARD; PRT; 530 AA.  
 AC P10614;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 51 (EC 1.14.14.-) (CYPL1) (P450-L1A1) (Sterol 14-alpha demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).  
 GN ERG11 OR CYP51 OR YHR007C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE#88111027; PubMed=3322742;  
 RA Kalb V.F., Woods C.W., Turi T.G., Dey C.R., Sutter T.R., Loper J.C.;  
 RT "Primary structure of the P450 lanosterol demethylase gene from  
 RT Saccharomyces cerevisiae.";  
 RT DNA 6:529-537 (1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE#88326319; PubMed=3046615;  
 RA Ishida N., Aoyama Y., Hatanaka K., Oyama Y., Imajo S., Ishiguro M.,  
 RA Oshida T., Nakazato H., Noguchi T., Maitra U.S., Mohan V.P.,  
 RA Sprinson D.B., Yoshida Y.;  
 RT "A single amino acid substitution converts cytochrome P450 (14DM) to  
 RT an inactive form, cytochrome P450G1: complete primary structures  
 RT deduced from cloned DNAs.";  
 RT Biochem. Biophys. Res. Commun. 155:317-323 (1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE#94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
 RA Nham M., Rifkin L., Riles L., St Peter H., Trevaekis E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
 RA Vaudin M.;  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT VII.";  
 RT Science 265:2077-2082 (1994).  
 RN [4]  
 RP SEQUENCE OF 444-530 FROM N.A.  
 RX MEDLINE#87106820; PubMed=3542713;  
 RA Kalb V.F., Loper J.C., Dey C.R., Woods C.W., Sutter T.R.;  
 RT "Isolation of a cytochrome P-450 structural gene from Saccharomyces  
 RT cerevisiae.";  
 RT Gene 45:237-245 (1986).  
 CC -I- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL WHICH IS  
 CC CRITICAL FOR ERGOSTEROL BIOSYNTHESIS. IT TRANSFORMS LANOSTEROL  
 CC INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL.  
 CC -I- PATHWAY: Ergosterol biosynthesis.  
 CC -I- MISCELLANEOUS: IT IS THE MAIN TARGET FOR ANTIFUNGAL COMPOUNDS OF  
 CC THE TRIAZOLE FAMILY LIKE KETOCONAZOLE WHICH INHIBITS BY  
 CC COORDINATING THE IRON ATOM AT THE SIXTH LIGAND POSITION.

CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M18109; AAA34379.1; -  
 DR EMBL; M15663; AAA34837.1; -  
 DR EMBL; M21483; AAA34546.1; -  
 DR EMBL; M21484; AAA34547.1; -  
 DR EMBL; U10555; AAB68433.1; -  
 DR PIR; A25563; A25563.  
 DR PIR; A27491; A27491.  
 DR PIR; B31569; B31569.  
 DR SGD; S0001049; ERG11.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450.1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450.1.  
 KM Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;  
 KW Sterol biosynthesis.  
 FT BINDING 470 HEME (BY SIMILARITY).  
 FT CONFLICT 433 K -> N (IN REF. 2).  
 FT SEQUENCE 530 AA; 60720 MW; 646960BBA0E17979 CRC64;

Query Match 51.9%; Score 41; DB 1; Length 530;  
 Best Local Similarity 61.5%; Pred. No. 5.7;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 SKNPHLOXNTIG 14  
 181 SKNPHLOXNTIG 193

RESULT 7  
 ID\_Y364\_RICPR STANDARD; PRT; 306 AA.  
 AC Q92DG3;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein RP364.  
 GN RP364.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OC NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E;  
 RX MEDLINE#99039499; PubMed=9823893;  
 RA Anderson S.G.E., Zomrodipour A., Andersson J.O.,  
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RT Nature 396:133-140 (1998).  
 RN [2]  
 RP -I- SIMILARITY: SOME, TO R. PROWAZEKII RP363.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ235271; CAA14823.1; -  
 DR Hypothetical protein; Complete proteome.



FT ACT\_SITE 521 521 GENERAL ACID CATALYST (BY SIMILARITY).  
 FT ACT\_SITE 522 522 INTERACT WITH SUBSTRATES (BY SIMILARITY).  
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 767 AA; 82488 MW; A5F29E2427EDB593 CRC64;  
 Query Match 50.6%; Score 40; DB 1; Length 767;  
 Best Local Similarity 57.1%; Pred. No. 13;  
 Matches 8; Conservative 3; Mismatches 1; Indels 2; Gaps 1;  
 QY 4 NEHOK-NTIGTG 15  
 :|||:|:|:  
 Db 416 SFHLQRTNNTLGGG 429  
 RESULT 10  
 ID AMY1\_SACDI STANDARD; PRT; 768 AA.  
 AC P29760;  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glucoamylase S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase) (GAL1).  
 GN ST42 OR DEX1.  
 OS Saccharomyces diastaticus (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=41870;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94288586; PubMed=8017901;  
 RA Kim K., Bajszar G., Lee S.Y., Knudsen F., Mattoon J.R.;  
 RT "Cloning of a new allelic variant of a Saccharomycetales diastaticus  
 RT glucoamylase gene and its introduction into industrial yeasts.";  
 RL Appl. Biochem. Biotechnol. 44:161-185(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91276266; PubMed=2055484;  
 RA Lambrechts M.G., Pretorius I.S., Sollietti P., Marmur J.;  
 RT "Primary structure and regulation of a glucoamylase-encoding gene  
 RT (ST42) in Saccharomyces diastaticus.";  
 RL Gene 100:95-103(1991).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M60650; AAA35107.1; -  
 CC EMBL; M90490; AAA20560.1; -  
 CC PIR; S35895; S35895.  
 CC HSSP; P08017; IAYX.  
 CC InterPro; IPR000165; GH\_15.  
 CC Pfam; PF00723; Glyco\_hydro\_15; 1.  
 CC PROSITE; PS00820; GLUCOAMYLASE; 1.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 22 768  
 FT DOMAIN 22 348  
 FT DOMAIN 349 692  
 FT DOMAIN 693 768  
 FT BINDING 456 456  
 FT ACT\_SITE 519 519  
 FT ACT\_SITE 522 522  
 FT ACT\_SITE 523 523  
 FT CARBOHYD 309 309  
 FT CARBOHYD 323 323  
 FT CARBOHYD 415 415  
 FT CARBOHYD 424 424  
 FT CARBOHYD 435 435  
 FT CARBOHYD 514 514  
 FT CARBOHYD 547 547  
 FT CARBOHYD 646 646  
 FT CARBOHYD 651 651  
 FT CARBOHYD 721 721  
 FT CARBOHYD 742 742  
 FT CONFLICT 164 164  
 FT CONFLICT 624 624  
 SQ SEQUENCE 768 AA; 82586 MW; 3FAC172C128A0C6F CRC64;  
 Query Match 50.6%; Score 40; DB 1; Length 768;  
 Best Local Similarity 57.1%; Pred. No. 13;  
 Matches 8; Conservative 3; Mismatches 1; Indels 2; Gaps 1;  
 QY 4 NEHOK-NTIGTG 15  
 :|||:|:|:  
 Db 417 SFHLQRTNNTLGGG 430  
 RESULT 11  
 ID A29B\_DROME STANDARD; PRT; 234 AA.  
 AC 046157; Q9TWT3; Q9TWT5; Q9TWT6; Q9TWT7; Q9U976; Q9U977; Q9U978;  
 AC Q9U979; Q9V305;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Accessory gland protein Acp29Ab precursor.  
 GN ACP29AB OR CG1797.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=Canton-S; TISSUE=Male accessory gland;  
 RX MEDLINE=98135120; PubMed=9474779;  
 RA Wolfner M.F., Harada H.A., Berttram M.J., Steelick T.J., Kraus K.W.,  
 RA Kalb J.M., Lung Y.O., Neubaum D.M., Park M., Tram U.K.;  
 RT "New genes for male accessory gland proteins in Drosophila  
 RT melanogaster.";  
 RL Insect Biochem. Mol. Biol. 27:825-834(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Various strains;  
 RX MEDLINE=99282496; PubMed=10353896;  
 RA Aguade M.;  
 RT "Positive selection drives the evolution of the Acp29Ab accessory  
 RT gland protein in Drosophila.";  
 RL Genetics 152:543-551(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Various strains;  
 RX MEDLINE=2056153; PubMed=1102381;  
 RA Begun D.J., Whitely P., Todd B.L., Waldrip-Dail H.M., Clark A.G.;

"Molecular population genetics of male accessory gland proteins in *Drosophila*.";  
 Genetics 156:1879-1888 (2000).  
 [4]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN-Berkeley;  
 MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers J.H.C., Blazer V.G., Chape M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abail J.F., Agbayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Daveport L.B., Davies P.,  
 de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Dup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mohr C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pauley J.M.,  
 Palazzolo M., Peltan G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weisscock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 "The genome sequence of *Drosophila melanogaster*."  
 Science 287:2185-2195(2000).  
 RT Science 287:2185-2195(2000).  
 CC -1- FUNCTION: RESPONSIBLE FOR PHYSIOLOGICAL AND BEHAVIORAL CHANGES IN  
 MALE FEMALE FLIES.  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLAND AND IN  
 SEMINAL FLUID.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by, and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL, U85758; AAB96382.1; -  
 DR EMBL, AJ240513; CAB53187.1; -  
 DR EMBL, AJ240514; CAB53188.1; -  
 DR EMBL, AJ240515; CAB53189.1; -  
 DR EMBL, AJ240516; CAB53190.1; -  
 DR EMBL, AJ240517; CAB53191.1; -  
 DR EMBL, AJ240518; CAB53192.1; -  
 DR EMBL, AJ240519; CAB53193.1; -  
 DR EMBL, AJ240520; CAB53194.1; -  
 DR EMBL, AJ240521; CAB53195.1; -  
 DR EMBL, AJ240522; CAB53196.1; -  
 DR EMBL, AJ240523; CAB53197.1; -  
 DR EMBL, AJ240524; CAB53198.1; -

DR EMBL, AJ240525; CAB53199.1; -  
 DR EMBL, AJ240526; CAB53200.1; -  
 DR EMBL, AJ240527; CAB53201.1; -  
 DR EMBL, AJ240528; CAB53202.1; -  
 DR EMBL, AJ240529; CAB53203.1; -  
 DR EMBL, AJ240530; CAB53204.1; -  
 DR EMBL, AJ240531; CAB53205.1; -  
 DR EMBL, AJ240532; CAB53206.1; -  
 DR EMBL, AJ240533; CAB53207.1; -  
 DR EMBL, AJ240534; CAB53208.1; -  
 DR EMBL, AJ240535; CAB53209.1; -  
 DR EMBL, AJ240536; CAB53210.1; -  
 DR EMBL, AJ240537; CAB53211.1; -  
 DR EMBL, AJ240538; CAB53212.1; -  
 DR EMBL, AJ240539; CAB53213.1; -  
 DR EMBL, AJ240540; CAB53214.1; -  
 DR EMBL, AJ240541; CAB53215.1; -  
 DR EMBL, AJ240542; CAB53216.1; -  
 DR EMBL, AJ240543; CAB53217.1; -  
 DR EMBL, AJ240544; CAB53218.1; -  
 DR EMBL, AJ240545; CAB53219.1; -  
 DR EMBL, AJ240546; CAB53220.1; -  
 DR EMBL, AJ240547; CAB53221.1; -  
 DR EMBL, AJ240548; CAB53222.1; -  
 DR EMBL, AJ240549; CAB53223.1; -  
 DR EMBL, AJ240550; CAB53224.1; -  
 DR EMBL, AJ240551; CAB53225.1; -  
 DR EMBL, AJ240552; CAB53259.1; -  
 DR EMBL, AV010527; AAG32559.1; -  
 DR EMBL, AV010528; AAG32560.1; -  
 DR EMBL, AV010529; AAG32561.1; -  
 DR EMBL, AV010530; AAG32562.1; -  
 DR EMBL, AV010531; AAG32563.1; -  
 DR EMBL, AV010532; AAG32564.1; -  
 DR EMBL, AV010533; AAG32565.1; -  
 DR EMBL, AV010534; AAG32566.1; -  
 DR EMBL, AV010535; AAG32567.1; -  
 DR EMBL, AV010536; AAG32568.1; -  
 DR EMBL, AV010537; AAG32569.1; -  
 DR EMBL, AV010538; AAG32570.1; -  
 DR EMBL, AV010539; AAG32571.1; -  
 DR EMBL, AV010540; AAG32572.1; -  
 DR EMBL, AV010541; AAG32573.1; -  
 DR EMBL, AV010542; AAG32574.1; -  
 DR EMBL, AV010543; AAG32575.1; -  
 DR EMBL, AV010544; AAG32576.1; -  
 DR EMBL, AV010545; AAG32577.1; -  
 DR EMBL, AV010546; AAG32578.1; -  
 DR EMBL, AV010547; AAG32579.1; -  
 DR EMBL, AV010548; AAG32580.1; -  
 DR EMBL, AV010549; AAG32581.1; -  
 DR EMBL, AV010550; AAG32582.1; -  
 DR EMBL, AV010551; AAG32583.1; -  
 DR EMBL, AV010552; AAG32584.1; -  
 DR EMBL, AV010553; AAG32585.1; -  
 DR EMBL, AV010554; AAG32586.1; -  
 DR EMBL, AV010555; AAG32587.1; -  
 DR EMBL, AV010556; AAG32588.1; -  
 DR EMBL, AV010557; AAG32589.1; -  
 DR EMBL, AV010558; AAG32590.1; -  
 DR EMBL, AV010559; AAG32591.1; -  
 DR EMBL, AV010560; AAG32592.1; -  
 DR EMBL, AV010561; AAG32593.1; -  
 DR EMBL, AV010562; AAG32594.1; -  
 DR EMBL, AV010563; AAG32595.1; -  
 DR EMBL, AV010564; AAG32596.1; -  
 DR EMBL, AV010565; AAG32597.1; -  
 DR EMBL, AV010566; AAG32598.1; -  
 DR EMBL, AV010567; AAG32599.1; -  
 DR EMBL, AV010568; AAG32600.1; -  
 DR EMBL, AV010569; AAG32601.1; -  
 DR EMBL, AV010570; AAG32602.1; -  
 DR EMBL, AV010571; AAG32603.1; -  
 DR EMBL, AV010572; AAG32604.1; -  
 DR EMBL, AV010573; AAG32605.1; -  
 DR EMBL, AV010574; AAG32606.1; -  
 DR EMBL, AV010575; AAG32607.1; -  
 DR EMBL, AV010576; AAG32608.1; -  
 DR EMBL, AV010577; AAG32609.1; -  
 DR EMBL, AV010578; AAG32610.1; -  
 DR EMBL, AV010579; AAG32611.1; -  
 DR EMBL, AV010580; AAG32612.1; -  
 DR EMBL, AV010581; AAG32613.1; -  
 DR EMBL, AV010582; AAG32614.1; -  
 DR EMBL, AV010583; AAG32615.1; -  
 DR EMBL, AV010584; AAG32616.1; -  
 DR EMBL, AV010585; AAG32617.1; -  
 DR EMBL, AV010586; AAG32618.1; -  
 DR EMBL, AV010587; AAG32619.1; -  
 DR EMBL, AV010588; AAG32620.1; -  
 DR EMBL, AV010589; AAG32621.1; -  
 DR EMBL, AV010590; AAG32622.1; -  
 DR EMBL, AV010591; AAG32623.1; -  
 DR EMBL, AV010592; AAG32624.1; -  
 DR EMBL, AV010593; AAG32625.1; -  
 DR EMBL, AV010594; AAG32626.1; -  
 DR EMBL, AV010595; AAG32627.1; -  
 DR EMBL, AV010596; AAG32628.1; -  
 DR EMBL, AV010597; AAG32629.1; -  
 DR EMBL, AV010598; AAG32630.1; -  
 DR EMBL, AV010599; AAG32631.1; -  
 DR EMBL, AV010600; AAG32632.1; -  
 DR EMBL, AV010601; AAG32633.1; -  
 DR EMBL, AV010602; AAG32634.1; -  
 DR EMBL, AV010603; AAG32635.1; -  
 DR EMBL, AV010604; AAG32636.1; -  
 DR EMBL, AV010605; AAG32637.1; -  
 DR EMBL, AV010606; AAG32638.1; -  
 DR EMBL, AV010607; AAG32639.1; -  
 DR EMBL, AV010608; AAG32640.1; -  
 DR EMBL, AV010609; AAG32641.1; -  
 DR EMBL, AV010610; AAG32642.1; -  
 DR EMBL, AV010611; AAG32643.1; -  
 DR EMBL, AV010612; AAG32644.1; -  
 DR EMBL, AV010613; AAG32645.1; -  
 DR EMBL, AV010614; AAG32646.1; -  
 DR EMBL, AV010615; AAG32647.1; -  
 DR EMBL, AV010616; AAG32648.1; -  
 DR EMBL, AV010617; AAG32649.1; -  
 DR EMBL, AV010618; AAG32650.1; -  
 DR EMBL, AV010619; AAG32651.1; -  
 DR EMBL, AV010620; AAG32652.1; -  
 DR EMBL, AV010621; AAG32653.1; -  
 DR EMBL, AV010622; AAG32654.1; -  
 DR EMBL, AV010623; AAG32655.1; -  
 DR EMBL, AV010624; AAG32656.1; -  
 DR EMBL, AV010625; AAG32657.1; -  
 DR EMBL, AV010626; AAG32658.1; -  
 DR EMBL, AV010627; AAG32659.1; -  
 DR EMBL, AV010628; AAG32660.1; -  
 DR EMBL, AV010629; AAG32661.1; -  
 DR EMBL, AV010630; AAG32662.1; -  
 DR EMBL, AV010631; AAG32663.1; -  
 DR EMBL, AV010632; AAG32664.1; -  
 DR EMBL, AV010633; AAG32665.1; -  
 DR EMBL, AV010634; AAG32666.1; -  
 DR EMBL, AV010635; AAG32667.1; -  
 DR EMBL, AV010636; AAG32668.1; -  
 DR EMBL, AV010637; AAG32669.1; -  
 DR EMBL, AV010638; AAG32670.1; -  
 DR EMBL, AV010639; AAG32671.1; -  
 DR EMBL, AV010640; AAG32672.1; -  
 DR EMBL, AV010641; AAG32673.1; -  
 DR EMBL, AV010642; AAG32674.1; -  
 DR EMBL, AV010643; AAG32675.1; -  
 DR EMBL, AV010644; AAG32676.1; -  
 DR EMBL, AV010645; AAG32677.1; -  
 DR EMBL, AV010646; AAG32678.1; -  
 DR EMBL, AV010647; AAG32679.1; -  
 DR EMBL, AV010648; AAG32680.1; -  
 DR EMBL, AV010649; AAG32681.1; -  
 DR EMBL, AV010650; AAG32682.1; -  
 DR EMBL, AV010651; AAG32683.1; -  
 DR EMBL, AV010652; AAG32684.1; -  
 DR EMBL, AV010653; AAG32685.1; -  
 DR EMBL, AV010654; AAG32686.1; -  
 DR EMBL, AV010655; AAG32687.1; -  
 DR EMBL, AV010656; AAG32688.1; -  
 DR EMBL, AV010657; AAG32689.1; -  
 DR EMBL, AV010658; AAG32690.1; -  
 DR EMBL, AV010659; AAG32691.1; -  
 DR EMBL, AV010660; AAG32692.1; -  
 DR EMBL, AV010661; AAG32693.1; -  
 DR EMBL, AV010662; AAG32694.1; -  
 DR EMBL, AV010663; AAG32695.1; -  
 DR EMBL, AV010664; AAG32696.1; -  
 DR EMBL, AV010665; AAG32697.1; -  
 DR EMBL, AV010666; AAG32698.1; -  
 DR EMBL, AV010667; AAG32699.1; -  
 DR EMBL, AV010668; AAG32700.1; -  
 DR EMBL, AV010669; AAG32701.1; -  
 DR EMBL, AV010670; AAG32702.1; -  
 DR EMBL, AV010671; AAG32703.1; -  
 DR EMBL, AV010672; AAG32704.1; -  
 DR EMBL, AV010673; AAG32705.1; -  
 DR EMBL, AV010674; AAG32706.1; -  
 DR EMBL, AV010675; AAG32707.1; -  
 DR EMBL, AV010676; AAG32708.1; -  
 DR EMBL, AV010677; AAG32709.1; -  
 DR EMBL, AV010678; AAG32710.1; -  
 DR EMBL, AV010679; AAG32711.1; -  
 DR EMBL, AV010680; AAG32712.1; -  
 DR EMBL, AV010681; AAG32713.1; -  
 DR EMBL, AV010682; AAG32714.1; -  
 DR EMBL, AV010683; AAG32715.1; -  
 DR EMBL, AV010684; AAG32716.1; -  
 DR EMBL, AV010685; AAG32717.1; -  
 DR EMBL, AV010686; AAG32718.1; -  
 DR EMBL, AV010687; AAG32719.1; -  
 DR EMBL, AV010688; AAG32720.1; -  
 DR EMBL, AV010689; AAG32721.1; -  
 DR EMBL, AV010690; AAG32722.1; -  
 DR EMBL, AV010691; AAG32723.1; -  
 DR EMBL, AV010692; AAG32724.1; -  
 DR EMBL, AV010693; AAG32725.1; -  
 DR EMBL, AV010694; AAG32726.1; -  
 DR EMBL, AV010695; AAG32727.1; -  
 DR EMBL, AV010696; AAG32728.1; -  
 DR EMBL, AV010697; AAG32729.1; -  
 DR EMBL, AV010698; AAG32730.1; -  
 DR EMBL, AV010699; AAG32731.1; -  
 DR EMBL, AV010700; AAG32732.1; -  
 DR EMBL, AV010701; AAG32733.1; -  
 DR EMBL, AV010702; AAG32734.1; -  
 DR EMBL, AV010703; AAG32735.1; -  
 DR EMBL, AV010704; AAG32736.1; -  
 DR EMBL, AV010705; AAG32737.1; -  
 DR EMBL, AV010706; AAG32738.1; -  
 DR EMBL, AV010707; AAG32739.1; -  
 DR EMBL, AV010708; AAG32740.1; -  
 DR EMBL, AV010709; AAG32741.1; -  
 DR EMBL, AV010710; AAG32742.1; -  
 DR EMBL, AV010711; AAG32743.1; -  
 DR EMBL, AV010712; AAG32744.1; -  
 DR EMBL, AV010713; AAG32745.1; -  
 DR EMBL, AV010714; AAG32746.1; -  
 DR EMBL, AV010715; AAG32747.1; -  
 DR EMBL, AV010716; AAG32748.1; -  
 DR EMBL, AV010717; AAG32749.1; -  
 DR EMBL, AV010718; AAG32750.1; -  
 DR EMBL, AV010719; AAG32751.1; -  
 DR EMBL, AV010720; AAG32752.1; -  
 DR EMBL, AV010721; AAG32753.1; -  
 DR EMBL, AV010722; AAG32754.1; -  
 DR EMBL, AV010723; AAG32755.1; -  
 DR EMBL, AV010724; AAG32756.1; -  
 DR EMBL, AV010725; AAG32757.1; -  
 DR EMBL, AV010726; AAG32758.1; -  
 DR EMBL, AV010727; AAG32759.1; -  
 DR EMBL, AV010728; AAG32760.1; -  
 DR EMBL, AV010729; AAG32761.1; -  
 DR EMBL, AV010730; AAG32762.1; -  
 DR EMBL, AV010731; AAG32763.1; -  
 DR EMBL, AV010732; AAG32764.1; -  
 DR EMBL, AV010733; AAG32765.1; -  
 DR EMBL, AV010734; AAG32766.1; -  
 DR EMBL, AV010735; AAG32767.1; -  
 DR EMBL, AV010736; AAG32768.1; -  
 DR EMBL, AV010737; AAG32769.1; -  
 DR EMBL, AV010738; AAG32770.1; -  
 DR EMBL, AV010739; AAG32771.1; -  
 DR EMBL, AV010740; AAG32772.1; -  
 DR EMBL, AV010741; AAG32773.1; -  
 DR EMBL, AV010742; AAG32774.1; -  
 DR EMBL, AV010743; AAG32775.1; -  
 DR EMBL, AV010744; AAG32776.1; -  
 DR EMBL, AV010745; AAG32777.1; -  
 DR EMBL, AV010746; AAG32778.1; -  
 DR EMBL, AV010747; AAG32779.1; -  
 DR EMBL, AV010748; AAG32780.1; -  
 DR EMBL, AV010749; AAG32781.1; -  
 DR EMBL, AV010750; AAG32782.1; -  
 DR EMBL, AV010751; AAG32783.1; -  
 DR EMBL, AV010752; AAG32784.1; -  
 DR EMBL, AV010753; AAG32785.1; -  
 DR EMBL, AV010754; AAG32786.1; -  
 DR EMBL, AV010755; AAG32787.1; -  
 DR EMBL, AV010756; AAG32788.1; -  
 DR EMBL, AV010757; AAG32789.1; -  
 DR EMBL, AV010758; AAG32790.1; -  
 DR EMBL, AV010759; AAG32791.1; -  
 DR EMBL, AV010760; AAG32792.1; -  
 DR EMBL, AV010761; AAG32793.1; -  
 DR EMBL, AV010762; AAG32794.1; -  
 DR EMBL, AV010763; AAG32795.1; -  
 DR EMBL, AV010764; AAG32796.1; -  
 DR EMBL, AV010765; AAG32797.1; -  
 DR EMBL, AV010766; AAG32798.1; -  
 DR EMBL, AV010767; AAG32799.1; -  
 DR EMBL, AV010768; AAG32800.1; -  
 DR EMBL, AV010769; AAG32801.1; -  
 DR EMBL, AV010770; AAG32802.1; -  
 DR EMBL, AV010771; AAG32803.1; -  
 DR EMBL, AV010772; AAG32804.1; -  
 DR EMBL, AV010773; AAG32805.1; -  
 DR EMBL, AV010774; AAG32806.1; -  
 DR EMBL, AV010775; AAG32807.1; -  
 DR EMBL, AV010776; AAG32808.1; -  
 DR EMBL, AV010777; AAG32809.1; -  
 DR EMBL, AV010778; AAG32810.1; -  
 DR EMBL, AV010779; AAG32811.1; -  
 DR EMBL, AV010780; AAG32812.1; -  
 DR EMBL, AV010781; AAG32813.1; -  
 DR EMBL, AV010782; AAG32814.1; -  
 DR EMBL, AV010783; AAG32815.1; -  
 DR EMBL, AV010784; AAG32816.1; -  
 DR EMBL, AV010785; AAG32817.1; -  
 DR EMBL, AV010786; AAG32818.1; -  
 DR EMBL, AV010787; AAG32819.1; -  
 DR EMBL, AV010788; AAG32820.1; -  
 DR EMBL, AV010789; AAG32821.1; -  
 DR EMBL, AV010790; AAG32822.1; -  
 DR EMBL, AV010791; AAG32823.1; -  
 DR EMBL, AV010792; AAG32824.1; -  
 DR EMBL, AV010793; AAG32825.1; -  
 DR EMBL, AV010794; AAG32826.1; -  
 DR EMBL, AV010795; AAG32827.1; -  
 DR EMBL, AV010796; AAG32828.1; -  
 DR EMBL, AV010797; AAG32829.1; -  
 DR EMBL, AV010798; AAG32830.1; -  
 DR EMBL, AV010799; AAG32831.1; -  
 DR EMBL, AV010800; AAG32832.1; -  
 DR EMBL, AV010801; AAG32833.1; -  
 DR EMBL, AV010802; AAG32834.1; -  
 DR EMBL, AV010803; AAG32835.1; -  
 DR EMBL, AV010804; AAG32836.1; -  
 DR EMBL, AV010805; AAG32837.1; -  
 DR EMBL, AV010806; AAG32838.1; -  
 DR EMBL, AV010807; AAG32839.1; -  
 DR EMBL, AV010808; AAG32840.1; -  
 DR EMBL, AV010809; AAG32841.1; -  
 DR EMBL, AV010810; AAG32842.1; -  
 DR EMBL, AV010811; AAG32843.1; -  
 DR EMBL, AV010812; AAG32844.1; -  
 DR EMBL, AV010813; AAG32845.1; -  
 DR EMBL, AV010814; AAG32846.1; -  
 DR EMBL, AV010815; AAG32847.1; -  
 DR EMBL, AV010816; AAG32848.1; -  
 DR EMBL, AV010817; AAG32849.1; -  
 DR EMBL, AV010818; AAG32850.1; -  
 DR EMBL, AV010819; AAG32851.1; -  
 DR EMBL, AV010820; AAG32852.1; -  
 DR EMBL, AV010821; AAG32853.1; -  
 DR EMBL, AV010822; AAG32854.1; -  
 DR EMBL, AV010823; AAG32855.1; -  
 DR EMBL, AV010824; AAG32856.1; -  
 DR EMBL, AV010825; AAG32857.1; -  
 DR EMBL, AV010826; AAG32858.1; -  
 DR EMBL, AV010827; AAG32859.1; -  
 DR EMBL, AV010828; AAG32860.1; -  
 DR EMBL, AV010829; AAG32861.1; -  
 DR EMBL, AV010830; AAG32862.1; -  
 DR EMBL, AV010831; AAG32863.1; -  
 DR EMBL, AV010832; AAG32864.1; -  
 DR EMBL, AV010833; AAG32865.1; -  
 DR EMBL, AV010834; AAG32866.1; -  
 DR EMBL, AV010835; AAG32867.1; -  
 DR EMBL, AV010836; AAG32868.1; -  
 DR EMBL, AV010837; AAG32869.1; -  
 DR EMBL, AV010838; AAG32870.1; -  
 DR EMBL, AV010839; AAG32871.1; -  
 DR EMBL, AV010840; AAG32872.1; -  
 DR EMBL, AV010841; AAG32873.1; -  
 DR EMBL, AV010842; AAG32874.1; -  
 DR EMBL, AV010843; AAG32875.1; -  
 DR EMBL, AV010844; AAG32876.1; -  
 DR EMBL, AV010845; AAG32877.1; -  
 DR EMBL, AV010846; AAG32878.1; -  
 DR EMBL, AV010847; AAG32879.1; -  
 DR EMBL, AV010848; AAG32880.1; -  
 DR EMBL, AV010849; AAG32881.1; -  
 DR EMBL, AV010850; AAG32882.1; -  
 DR EMBL, AV010851; AAG32883.1; -  
 DR EMBL, AV010852; AAG32884.1; -  
 DR EMBL, AV010853; AAG32885.1; -  
 DR EMBL, AV010854; AAG32886.1; -  
 DR EMBL, AV010855; AAG32887.1; -  
 DR

FT MA5, MA21, MA45, MA52, MA67, ZIM29,  
ZIM30, ZIM42 AND ZIM56).  
E -> D (IN STRAIN MA7).  
SQ SEQUENCE 234 AA; 27173 MW; 8954CD3215480F3E CRC64;

Query Match 49.4%; Score 39; DB 1; Length 234;  
Best Local Similarity 46.2%; Pred. No. 5.5;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 SKNFHLQKNTI 14  
DB 119 SNFHLKNTI 131

## RESULT 12

QYEA STRPN STANDARD; PRT; 342 AA.

AC Q97015; 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE S-adenosylmethionine:RNA ribosyltransferase-isomerase (EC 5.-.-.-)  
GN QYEA OR SP1416.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=13113;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TIGR4;  
RX MEDLINE=21357209; PubMed=1163916;  
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayahara L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Lotz B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RU Science 293:498-506(2001).  
CC -1- FUNCTION: Synthesizes oq from preo1 in a single S-  
CC adenosylmethionine-requiring step. The ribosyl moiety of Adomet is  
CC transferred and isomerized to the epoxycyclopentane residue of oq  
CC (By similarity).  
CC -1- PATHWAY: Quenosine biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO THE QYEA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE007438; AAK75514.1; -.  
CC TIGR; SP1416; -.  
CC InterPro: IPR003699; Quenosine synth.  
CC Pfam: PF02547; Quenosine synth. 1.  
CC TrEMBL; TIGR00113; queA.1.  
CC DR TrEMBL; TIGR00113; queA.1.  
CC QY SEQUENCE 342 AA; 38379 MW; E81265A1C36D69 CRC64;

Query Match 49.4%; Score 39; DB 1; Length 342;  
Best Local Similarity 63.6%; Pred. No. 8.4;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 SKNFHLQKNTI 12

DB 295 STNFHLPKSTL 305

## RESULT 13

QYEA STRPY STANDARD; PRT; 342 AA.

AC Q99249; 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE S-adenosylmethionine:RNA ribosyltransferase-isomerase (EC 5.-.-.-)  
GN QYEA OR SPY1400.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SF370 / ATCC 700294 / Serotype M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti V.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RU Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
CC -1- FUNCTION: Synthesizes oq from preo1 in a single S-  
CC adenosylmethionine-requiring step. The ribosyl moiety of Adomet is  
CC transferred and isomerized to the epoxycyclopentane residue of oq  
CC (By similarity).  
CC -1- PATHWAY: Quenosine biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO THE QYEA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE006577; AAK34216.1; -.  
CC InterPro: IPR003699; Quenosine synth.  
CC Pfam: PF02547; Quenosine synth. 1.  
CC TrEMBL; TIGR00113; queA.1.  
CC DR TrEMBL; TIGR00113; queA.1.  
CC QY SEQUENCE 342 AA; 38351 MW; E46CD9203E3FDBF2 CRC64;

Query Match 49.4%; Score 39; DB 1; Length 342;  
Best Local Similarity 63.6%; Pred. No. 8.4;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 SKNFHLQKNTI 12  
DB 295 STNFHLPKSTL 305

## RESULT 14

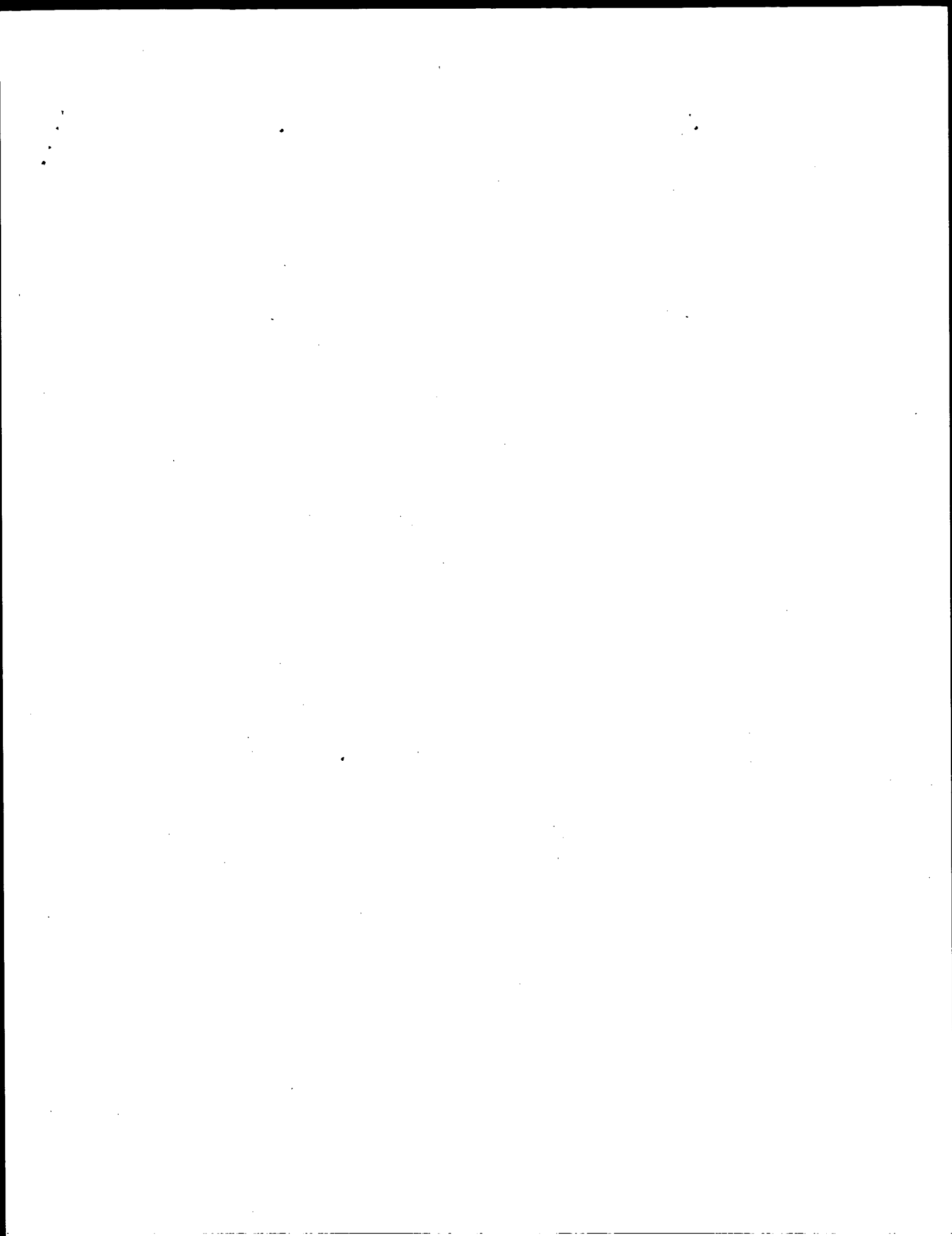
PEPG\_LACDL STANDARD; PRT; 437 AA.

AC P94869; 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Amino-peptidase G (EC 3.4.22.-).  
GN PEPG.  
OS Lactobacillus delbrueckii (subsp. lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=29397;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 7290;  
 RX MEDLINE=97195798; PubMed=9043129;  
 RA Klein J.R., Schick J., Henrich B., Plapp R.;  
 RT "necrobacillus delbrueckii subsp. lactis DSM7290 pepg gene encodes a  
 novel cysteine aminopeptidase.";  
 RL Microbiology 143:527-537(1997).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z71782; CA96465.1; -.  
 CC HSP: Q13867; ZCB5.  
 DR DR  
 DR MEROPS; C01.089; -.  
 DR InterPro: IPR004134; Pept C1 like.  
 DR InterPro: IPR000668; Peptidase C1.  
 DR InterPro: IPR000169; SHprot\_acsite.  
 DR Pfam: PF03051; Pept\_C1-like; 1.  
 DR PRINTS; PR00705; PAPAIN.  
 DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; 1.  
 DR PROSITE; PS00640; THIOL\_PROTEASE\_ASN; FALSE\_NEG.  
 KM Hydrolyase; Thiol protease; Aminopeptidase.  
 FT ACT\_SITE 70 70 BY SIMILARITY.  
 FT ACT\_SITE 361 361 BY SIMILARITY.  
 FT ACT\_SITE 382 382 BY SIMILARITY.  
 SQ SEQUENCE 437 AA; 49705 MW; 2099C62E57181FC5 CRC64;  
 Query Match 49.4%; Score 39; DB 1; Length 437;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KNFHLQKN 10  
 DB 234 KNVHLEKN 241  
 RESULT 15  
 FLIV SALMU STANDARD; PRT; 173 AA.  
 AC P37588;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Flagellar biosynthetic protein fliv.  
 GN FLIV.  
 OS Salmonella muenchen.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OK NCBI\_TaxID=596;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 8388;  
 RX MEDLINE=93231524; PubMed=8472952;  
 RA Doll L., Frankel G.;  
 RT "Cloning and sequencing of two new flv genes, the products of which  
 RT are essential for Salmonella flagellar biosynthesis.";  
 RL Gene 126:119-121(1993).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94075958; PubMed=8254311;  
 RA Doll L., Frankel G.;  
 RT "flv and fliv: two flagellar genes essential for biosynthesis of  
 RT Salmonella and Escherichia coli flagella.";  
 RL J. Gen. Microbiol. 139:2415-2422(1993).  
 CC -1- FUNCTION: Required for the secretion of flagellin and expression

CC of motility.  
 CC -1- SIMILARITY: BELONGS TO THE FLIV FAMILY. CORRESPONDS TO THE C-  
 CC TERMINAL SECTION  
 CC -1- CAUTION: Unlike the other Salmonellae, the ortholog of the flv  
 CC protein may be encoded by two CDS in S.muenchen. It cannot be  
 CC ruled out that sequencing errors produced two CDS instead of one.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L06521; AAA27107.1; -.  
 DR DR  
 DR PIR; JN0515; JN0515.  
 KW Flagella.  
 SQ SEQUENCE 173 AA; 19982 MW; 84DFAB36D1FF32C8 CRC64;  
 Query Match 48.1%; Score 38; DB 1; Length 173;  
 Best Local Similarity 50.0%; Pred. No. 6;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 SKNFHLQKNITGTV 15  
 DB 150 AKNFHRIETVATG 163

Search completed: April 20, 2003, 13:07:44  
 Job time : 3.92105 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07695 Seconds

(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524d-131

Sequence: 1 SRAEVSYPHNGAKF 15

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	100.0	514	2	S48730
2	76	100.0	514	2	JC2498
3	63	82.9	514	2	JC7100
4	61	80.3	507	2	JC7366
5	45	59.2	155	2	E97507
6	45	59.2	155	2	A12725
7	43	56.6	419	1	J00346
8	43	56.6	431	2	T46187
9	43	56.6	456	2	E69391
10	42	55.3	131	2	S42733
11	42	55.3	385	2	S53838
12	42	55.3	386	2	S58749
13	41	53.9	198	2	C90417
14	41	53.9	316	2	G72077
15	41	53.9	316	2	G86545
16	41	53.9	1303	1	S27396
17	40	52.6	256	2	S34245
18	40	52.6	374	2	T11323
19	40	52.6	389	2	S68155
20	40	52.6	578	2	B64012
21	39	51.3	362	2	S63538
22	39	51.3	386	1	S12352
23	39	51.3	576	2	AC2195
24	39	51.3	4930	2	C89447
25	38	50.0	89	2	C89447
26	38	50.0	97	2	H69895
27	38	50.0	343	2	S62704
28	38	50.0	348	2	A70311
29	38	50.0	385	1	CBBY

#### ALIGNMENTS

30	38	50.0	365	1	S15157	ubiquinol-cytochrome
31	38	50.0	365	2	S78660	ubiquinol-cytochrome
32	38	50.0	391	2	D89926	30S ribosomal prot
33	38	50.0	393	2	E82897	hypothetical prote
34	38	50.0	407	2	T19252	hypothetical prote
35	38	50.0	412	2	S23208	ubiquinol-cytochrome
36	38	50.0	423	1	MRBY	mRNA maturase b12
37	38	50.0	423	1	S78662	mRNA maturase b13
38	38	50.0	427	2	C97628	mRNA maturase SCB1
39	38	50.0	427	2	AE2851	ubiquinol-cytochrome
40	38	50.0	428	2	S75037	hypothetical prote
41	38	50.0	432	2	AD3311	folyl-polyglutamat
42	38	50.0	461	2	T08215	ubiquinol-cytochrome
43	38	50.0	517	2	S23209	polygalacturonase
44	38	50.0	517	2	S78664	mRNA maturase b13
45	38	50.0	589	2	G71325	mRNA maturase b13
						probable V-type AT

#### RESULT 1

S48730 Cry j II protein - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999

C/Accession: S48730

R/Namba, M.; Kurose, M.; Torioka, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.; FERS Lett. 353, 124-128, 1994

A/Title: Molecular cloning of the second major allergen, Cry j II, from Japanese cedar

A/Reference number: S48730, MUID:95010777, PMID:7926035

A/Accession: S48730

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-514 <KMW>

A/Cross-References: GB:D37765; NID:9577695; PIDN:BA07021.1; PID:dl007598; PID:g577694

Query Match

Best Local Similarity 100.0%; Score 76; DB 2; Length 514;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRAEVSYPHNGAKF 15

DB 290 SRAEVSYPHNGAKF 304

#### RESULT 2

JC2498 second major allergen Cry j II precursor - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)

C/Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000

C/Accession: JC2498; PC2346; A60147

R/Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994

A/Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese cedar

A/Reference number: JC2498; MUID:94271186; PMID:8002972

A/Accession: JC2498

A/Molecule type: mRNA

A/Residues: 1-514 <KMW>

A/Cross-References: DDBJ:D29772; NID:9506857; PIDN:BA06172.1; PID:g506858

A/Accession: PC2346

A/Molecule type: protein

A/Residues: 52-61 <KQ2>

R/Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matunasi, T.

Allergy 45, 309-312, 1990

A/Title: Identification of the second major allergen of Japanese cedar pollen.

A/Reference number: A60147; MUID:9034298; PMID:2382797

A/Accession: A60147

A/Molecule type: protein

A/Residues: 55-64 <SAK>

C/Keywords: glycoprotein; pollen

F:1-54/domain: signal sequence

F:55-460/Product: second major allergen Cry j

F:55-460/Product: second major allergen Cry j

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 76; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 8.2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRAEVSYVHNGAKF 15  
|||||  
Db 290 SRAEVSYVHNGAKF 304

## RESULT 3

JC7100 polygalacturonase Cha o 2 - Japanese cypress

C:Species: Chamaecyparis obtusa (Japanese cypress)

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

C:Accession: JC7100; PC7026

R:Morit, T.; Yokoyama, M.; Komiyama, N.; Okano, M.; Kino, K.

Biochem. Biophys. Res. Commun. 263, 166-171, 1999

A:Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A:Reference number: JC7100; PMID:99417540; PMID:10486272

A:Accession: JC7100

A:Molecule type: mRNA

A:Residues: 1-514 <MOR>

A:Accession: PC7026

A:Molecule type: protein

A:Residues: 51-62 <MO2>

Query Match 82.9%; Score 63; DB 2; Length 514;  
Best Local Similarity 86.7%; Pred. No. 0.0019;  
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SRAEVSYVHNGAKF 15  
|||||  
Db 290 SRAEVSYVHNGAKF 304

## RESULT 4

JC7366 Jun a 2 protein - mountain cedar

C:Species: Juniperus ashei (mountain cedar)

C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C:Accession: JC7366; PC7093

R:Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.

Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A:Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A:Reference number: JC7366

A:Accession: JC7366

A:Molecule type: mRNA

A:Residues: 1-507 <YOK>

A:Cross-references: GB:AJ04653

A:Accession: PC7093

A:Molecule type: protein

A:Residues: 55-63 <Y02>

C:Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

C:Keywords: glycoprotein; pollen

Query Match 80.3%; Score 61; DB 2; Length 507;  
Best Local Similarity 73.3%; Pred. No. 0.0044;  
Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRAEVSYVHNGAKF 15  
|||||  
Db 291 SRAEVSYVHNGAKF 305

## RESULT 5

E97507 hypothetical protein AGR\_C\_2237 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: E97507

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldma

A.; Liu, F.; Moliam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, E

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: E97507

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87014.1; PID:G15156260; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C\_2237

A:Map position: circular chromosome

Query Match 59.2%; Score 45; DB 2; Length 155;  
Best Local Similarity 61.5%; Pred. No. 0.98;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 RAEVSYVHNGAK 14  
|||||  
Db 132 RADVAIVHNSAR 144

## RESULT 6

A12725

conserved hypothetical protein Atu1212 [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: A12725

R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCle

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: A12725

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <KUR>

A:Cross-references: GB:AE008688; PIDN:AL42223.1; PID:G17739617; GSPDB:GN00186

C:Genetics:

A:Gene: Atu1212

A:Map position: circular chromosome

Query Match 59.2%; Score 45; DB 2; Length 155;  
Best Local Similarity 61.5%; Pred. No. 0.98;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 RAEVSYVHNGAK 14  
|||||  
Db 132 RADVAIVHNSAR 144

## RESULT 7

JQ0346

ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - Rhodospseudomonas viridi

C:Species: Rhodospseudomonas viridis

C:Date: 07-Sep-1990 #sequence\_revision 02-Aug-1994 #text\_change 03-Jun-2002

C:Accession: JQ0346

R:Verbit, J.; Lang, F.; Gabelin, N.; Oesterhel, D.

Mol. Gen. Genet. 219, 445-452, 1989

A:Title: Cloning and sequencing of the fbcF, B and C genes encoding the cytochrome b/h

A:Reference number: JQ0345; PMID:90158506; PMID:2560136

A:Accession: JQ0346

A:Molecule type: DNA

A:Residues: 1-419 <VER>

A:Experimental source: strain 133

C:Comment: This protein is one of the three subunits of the ubiquinol-cytochrome C2 c

lectron-transport system that catalyzes the synthesis of ATP.

C:Genetics:

A:Gene: fbcb  
 C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin  
 C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase  
 C:Accession: E69391  
 F:24-355/Domain: cytochrome b homology <CB6>  
 F:24-224/Domain: cytochrome b6 homology <CB6>  
 F:49-65/Domain: transmembrane #status predicted <TM1>  
 F:94-112/Domain: transmembrane #status predicted <TM2>  
 F:112-148/Domain: transmembrane #status predicted <TM3>  
 F:143-215/Domain: transmembrane #status predicted <TM4>  
 F:238-356/Domain: plastocyanin-plastocyanin reductase 17K protein homology <17K>  
 F:246-262/Domain: transmembrane #status predicted <TM5>  
 F:305-321/Domain: transmembrane #status predicted <TM6>  
 F:340-360/Domain: transmembrane #status predicted <TM7>  
 F:370-386/Domain: transmembrane #status predicted <TM8>  
 F:96-197/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted  
 F:110-211/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 56.6%; Score 43; DB 1; Length 419;  
 Best Local Similarity 54.5%; Pred. No. 6.8;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VSYVHNGAKF 15  
 Db 92 IRYIHNGASF 102

RESULT 8  
 T46187  
 polygalacturonase (EC 3.2.1.15) precursor [similarity] - Arabidopsis thaliana  
 N:Alternate names: protein T8H10.110  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Aug-2000  
 C:Accession: T46187; T50674  
 R:Genes, V.; Rechmann, S.; Borkova, D.; Ansoyge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23014  
 A:Accession: T46187  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-431 <BEN>  
 A:Cross-references: EMBL:AL133248; PIDN:CAB66108.1  
 A:Experimental source: cultivar Columbia; BAC clone T8H10  
 R:Jenkins, E.S.; Roberts, J.A.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: Dehiscence-related expression of an Arabidopsis thaliana gene encoding a  
 A:Reference number: Z25172  
 A:Accession: T50674  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-431 <JEN>  
 A:Cross-references: EMBL:AF037367; PIDN:AAC98923.1  
 A:Experimental source: cultivar Landsberg erecta  
 C:Genetics:  
 A:Gene: T8H10.110  
 A:Map position: 3  
 A:Introns: 85/3; 129/3; 186/3; 193/3; 263/1; 290/2; 326/3; 364/3  
 C:Superfamily: polygalacturonase  
 C:Keywords: glycosidase; hydrolase  
 F:1-27/Domain: signal sequence #status predicted <SIG>  
 F:28-431/Product: polygalacturonase #status predicted <MAT>

Query Match 56.6%; Score 43; DB 2; Length 431;  
 Best Local Similarity 60.0%; Pred. No. 7.1;  
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRRAVSIVHNGAKF 15  
 Db 299 SKAYVGINVDGAKF 313

RESULT 9  
 E69391  
 hypothetical protein AF1134 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: E69391  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E  
 Glodok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Moore, C.R.; Venter, J.C.  
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: E69391  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-456 <KLE>  
 A:Cross-references: GB:AE01026; GB:AE00782; NID:G2689349; PIDN:AB90123.1; PID:G264

Query Match 56.6%; Score 43; DB 2; Length 456;  
 Best Local Similarity 72.7%; Pred. No. 7.5;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 VSYVHNGAKF 15  
 Db 178 VSYVLDGAKF 188

RESULT 10  
 S42733  
 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - yeast (Williopsis mrakii  
 C:Keywords: chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative pho  
 C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 03-Jun-2002  
 C:Accession: S42733  
 R:Diriesi, R.; Sor, F.; Nosek, J.; Fukuhara, H.  
 Yeast 10, 391-398, 1994  
 A>Title: Genes of the linear mitochondrial DNA of Williopsis mrakii: coding sequences

A:Reference number: S42733; MUID:94287716; PMID:8017108  
 A:Accession: S42733  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-131 <DR1>  
 A:Cross-references: EMBL:X66594; NID:G296849; PIDN:CAA47157.1; PID:G296850  
 A:Experimental source: strain CBS1707  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992  
 C:Genetics:  
 A:Gene: CYTB  
 A:Genome: mitochondrion  
 A:Genetic code: SGC2  
 A:Introns: 67/3  
 C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin  
 C:Keywords: chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative pho  
 F:10-131/Domain: cytochrome b6 homology (fragment) <CB6>  
 F:110-131/Domain: cytochrome b homology (fragment) <CB6>  
 F:35-51/Domain: transmembrane #status predicted <TM1>  
 F:80-98/Domain: transmembrane #status predicted <TM2>

Query Match 55.3%; Score 42; DB 2; Length 131;  
 Best Local Similarity 54.5%; Pred. No. 2.9;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 VSYVHNGAKF 15  
 Db 78 IRYIHNGASF 88

RESULT 11  
 S53838  
 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - Acanthamoeba castellanii  
 C:Species: mitochondrion Acanthamoeba castellanii  
 C>Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 03-Jun-2002  
 C:Accession: S53838  
 R:Burger, G.; Plante, I.; Loneragan, K.M.; Gray, M.W.  
 J. Mol. Biol. 245, 522-537, 1995

A>Title: The mitochondrial DNA of the amoeboid protozoan, *Acanthamoeba castellanii*: comp  
 A:Reference number: S53825; MUID:95147275; PMID:7844823  
 A:Accession: S53838  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-385 <BUR>  
 A:Cross-references: GB:U12386; NID:9562028; PIDN:AA01830.1; PID:9562042  
 A:Experimental source: strain Neff; ATCC 30010  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
 C:Genetics:

A:Gene: cob  
 A:Genome: mitochondrion  
 A:Genetic code: SGC6  
 C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin  
 C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; mitochondrion;  
 F:10-341/Domain: cytochrome b homology <CBH>  
 F:10-210/Domain: cytochrome b6 homology <CB6>  
 F:35-51/Domain: transmembrane #status predicted <TM1>  
 F:80-98/Domain: transmembrane #status predicted <TM2>  
 F:118-134/Domain: transmembrane #status predicted <TM3>  
 F:179-201/Domain: transmembrane #status predicted <TM4>  
 F:223-341/Domain: plastocyanin reductase 17K protein homology <17K>  
 F:231-247/Domain: transmembrane #status predicted <TM5>  
 F:299-306/Domain: transmembrane #status predicted <TM6>  
 F:325-345/Domain: transmembrane #status predicted <TM7>  
 F:355-371/Domain: transmembrane #status predicted <TM8>  
 F:82-183/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted  
 F:96-197/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 55.3%; Score 42; DB 2; Length 385;  
 Best Local Similarity 54.5%; Pred. No. 9.5;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 VSYVHNGAKF 15  
 : : : : :  
 DB 78 IRYIHANGASF 88

RESULT 12  
 S58749  
 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - yeast (*Hansenula wingei*) m  
 C:Species: mitochondrion *Hansenula wingei*  
 C>Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 03-Jun-2002  
 C:Accession: S58749  
 R:Seikito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.  
 Curr. Genet. 28, 39-53, 1995  
 A>Title: The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new charact  
 A:Reference number: S58740; MUID:96022424; PMID:8536312  
 A:Accession: S58749  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-386 <SEK>  
 A:Cross-references: EMBL:D31785  
 C:Genetics:

A:Genome: mitochondrion  
 A:Genetic code: SGC3  
 C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin  
 C:Keywords: electron transfer; heme; iron; metalloprotein; mitochondrion; oxidative phos  
 F:10-340/Domain: cytochrome b homology <CYB>  
 F:10-210/Domain: cytochrome b6 homology <CB6>  
 F:222-340/Domain: plastocyanin reductase 17K protein homology <17K>  
 F:82-183/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted  
 F:96-197/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 55.3%; Score 42; DB 2; Length 386;  
 Best Local Similarity 54.5%; Pred. No. 9.5;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 VSYVHNGAKF 15  
 : : : : :  
 DB 78 IRYIHANGASF 88

RESULT 13  
 C90417  
 endonuclease V (nfi) [imported] - *Sulfolobus solfataricus*  
 C:Species: *Sulfolobus solfataricus*  
 C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
 C:Accession: C90417  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awaize, M.J.; Ch  
 Jung, I.; Ueffler, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
 arrett, R.A.; Reagan, W.A.; Jensen, C.W.; Van der Oost, J.  
 submitted to Genbank, April 2001  
 A:Description: *Sulfolobus solfataricus* complete genome.  
 A:Reference number: A99139  
 A:Accession: C90417  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-198 <RUR>  
 A:Cross-references: GB:AE00641; NID:G13815757; PIDN:AAK42594.1; GSPDB:GN00155  
 C:Genetics:

A:Superfamily: conserved hypothetical protein AF0129  
 Query Match 53.9%; Score 41; DB 2; Length 198;  
 Best Local Similarity 42.9%; Pred. No. 6.9;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRAEVSYVHNGAK 14  
 : : : : :  
 DB 139 NESITVYVINGEK 152

RESULT 14  
 G72077  
 ct007 hypothetical protein - *Chlamydomonas pneumoniae* (strain CML029)  
 C:Species: *Chlamydomonas pneumoniae*, *Chlamydia pneumoniae*  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
 C:Accession: G72077  
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
 Nature Genet. 21, 385-389, 1999  
 A>Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: G72077  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-316 <ARN>  
 A:Cross-references: GB:AE001627; GB:AE001363; NID:G4376721; PIDN:AA018585.1; PID:G437  
 A:Experimental source: strain CML029  
 C:Genetics:

Query Match 53.9%; Score 41; DB 2; Length 316;  
 Best Local Similarity 61.5%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AEVSYVHNGAKF 15  
 : : : : :  
 DB 44 AQVQYLKVNDAKF 56

RESULT 15  
 G86545  
 CT007 hypothetical protein [imported] - *Chlamydomonas pneumoniae* (strain J138)  
 C:Species: *Chlamydomonas pneumoniae*, *Chlamydia pneumoniae*  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: G86545  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A>Title: Comparison of whole genome sequences of *Chlamydia pneumoniae* J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: G86545  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-316 <STO>  
 A:Cross-references: GB:BA000008; NID:G8978813; PIDN:BA098649.1; GSPDB:GN00142

Sat May 10 '08:43:40 2003

us-09-142-524d-131.rpr

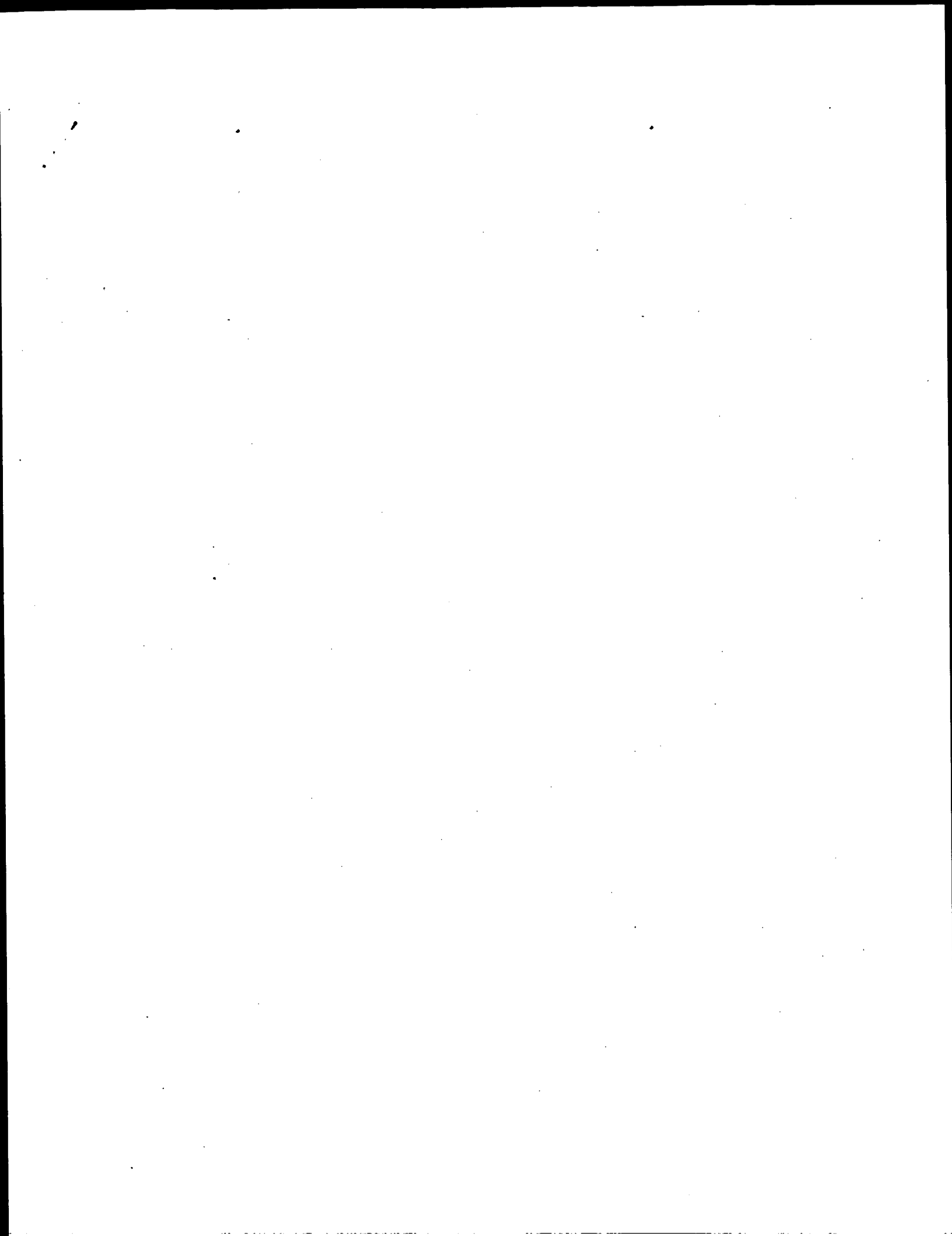
Page 5

A/Experimental source: strain J138  
C/Genetics:  
A/Gene: CPj0441

Query Match	53.9%	Score 41;	DB 2;	Length 31c;
Best Local Similarity	61.5%	Pred. NO. 12;		
Matches	8;	Conservative	2;	Mismatches 3;
				Indels 0;
				Gaps 0;

Qy	3	AEVSYVHNGAKF	15
		:     :	
Db	4 4	AQVQYLKVNDAKF	56

Search completed: April 20, 2003, 13:15:57  
Job time : 8.07895 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 Seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-131

Sequence: 1 SRAEVSYHVNAGKF 15

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	100.0	514	MPA2_CRYJA	P43212 cryptomeria
2	43	56.6	419	CYB_RHOVI	P81378 rhodopseudo
3	42	55.3	385	CYB_ACACA	Q37378 acanthamoeb
4	42	55.3	386	CYB_PICPT	Q36507 pichia plip
5	41	53.9	316	V441_CHEPU	Q36842 chlamydia p
6	41	53.9	1307	PHYI_CERPU	P25848 ceratodon p
7	40	52.6	379	CYB_IOLBL	O47477 loligo blae
8	40	52.6	389	CYB_DICDI	Q37311 dictyosteli
9	40	52.6	578	VTFM_HAEIN	P44038 haemophilus
10	39	51.3	467	PGLR_ACRCH	P35336 acclimidia c
11	38	50.0	134	CYB_GORDA	Q21411 sorex dapha
12	38	50.0	348	V113_AQUAE	O66516 aquifex aeo
13	38	50.0	385	CYB_SACDO	Q35819 saccharomyc
14	38	50.0	385	CYB_YEAST	P00163 saccharomyc
15	38	50.0	386	CYB_HANWI	P48877 hansenula w
16	38	50.0	423	MBI2_YEAST	P02873 saccharomyc
17	38	50.0	589	VAAL1_TREPA	Q83441 treponema p
18	38	50.0	589	VAAL1_TREPA	Q83441 treponema p
19	38	50.0	589	VAAL1_TREPA	Q83441 treponema p
20	37	48.7	341	MURB_PASMU	P57952 pasteurella
21	37	48.7	357	TRMA_CAMJE	Q99p92 campylobact
22	37	48.7	379	CYB_URSAR	Q36192 ursus arcto
23	37	48.7	379	CYB_URSMA	Q36082 ursus marit
24	37	48.7	380	CYB_ONCMW	Q45173 oncomychnu
25	37	48.7	380	CYB_SALSA	Q35925 salmo salar
26	37	48.7	380	CYB_SALSA	Q35925 salmo salar
27	37	48.7	383	CYB_PHYME	P29671 phytophthor
28	37	48.7	399	ZOT_VIBCH	P38442 vibrio choi
29	37	48.7	607	MODM_RHIVV	P08633 rhizobium l
30	37	48.7	687	XKDV_BACSU	P53431 bacillus su
31	37	48.7	733	AGAI_PEDPE	P43467 pediococcus
32	37	48.7	1385	NAC2_CHURE	Q91em8 chlamydomon
33	37	48.7	2111	YPB4_CABEL	Q11107 caenorhabdi

## ALIGNMENTS

RESULT 1	MPA2_CRYJA	STANDARD:	PRT:	514 AA.	
AC	P43212;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Possible polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)				
DE	(Major pollen allergen Cry j 2) (Cry j II).				
OS	Cryptomeria japonica (Japanese cedar).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.				
OX	NCBI_TaxId=3369;				
RN	(1)				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Pollen;				
RX	MEDLINE=9501077; PubMed=7926035;				
RA	Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.,				
RT	Usui M., Kurimoto M.,				
RT	"Molecular cloning of the second major allergen, Cry j II, from				
RT	Japanese cedar pollen.";				
RL	FEBS Lett. 353:124-128 (1994).				
RN	(2)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pollen;				
RX	MEDLINE=94271186; PubMed=8002972;				
RA	Komiyama N., Sone T., Shimizu K., Morikubo K., Kino K.;				
RT	"CDNA cloning and expression of Cry j II the second major allergen of				
RT	Japanese cedar pollen.";				
RL	Biochem. Biophys. Res. Commun. 201:1021-1028(1994).				
RN	(3)				
RP	SEQUENCE OF 55-64.				
RX	MEDLINE=90342988; PubMed=2382797;				
RA	Sakaguchi M., Inouye S., Tanai T., Ando S., Usui M., Matubasi T.;				
RT	"Identification of the second major allergen of Japanese cedar				
RT	pollen.";				
RL	Allergy 45:309-312(1990).				
CC	-1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-				
CC	galactosiduronic linkages in pectate and other galacturonans.				
CC	-1- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).				
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES				
CC	(POLYGALACTURONASES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; D37765; BAA07021.1; -				
DR	EMBL; D29772; BAA06172.1; -				
DR	HSSP; P26509; 1BHE.				
DR	InterPro; IPR000743; GH28.				
DR	Pfam; PF00295; Glyco_hydro_28; 1.				

007575 bacillus su  
P13358 dipodomys h  
P15357 dipodomys p  
Q56206 s glucosami  
Q35131 nycticebus  
Q9xp68 pseudoryx n  
Q96693 romerolagus  
P96154 vibrio choi  
O86648 a glucosami  
Q9abv2 c glucosami  
O8ueh1 a glucosami  
Q92zk3 rhizobium m





CYB\_PICPJ STANDARD; PRT; 386 AA.  
 ID CYB\_PICPJ  
 AC Q36507;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Cytochrome B.  
 GN COB OR CYTB.  
 OS Pichia pilipeti (Yeast).  
 OG Mitochondrion.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Pichia.  
 OX NCBI\_TaxID=4928;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 2887;  
 RX MEDLINE=93204976; PubMed=8455612;  
 RA Fukushima H., Sor F., Drissi R., Dinouel N., Miyakawa I., Rousset S.,  
 RA Viola A.M.,  
 RT "Linear mitochondrial DNAs of yeasts: frequency of occurrence and  
 RT general features."  
 RL Mol. Cell. Biol. 13:2309-2314(1993).  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS.  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN.  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X66593; CAA47155.1; -  
 DR InterPro; IPR00179; Cyt b b6.  
 DR Pfam; PF00032; cytochrome\_b\_c1.  
 DR Pfam; PF00033; cytochrome\_b\_n; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 DR PROSITE; PS00193; CYTOCHROME\_B\_OO; FALSE NEG.  
 KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;  
 HM Heme.  
 FT METAL 82 82 IRON 1 (HEME B562 AXIAL LIGAND).  
 FT METAL 96 96 IRON 2 (HEME B566 AXIAL LIGAND).  
 FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).  
 FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).  
 SQ SEQUENCE 386 AA; 43345 MW; CS3A87644963E5B CRC64;  
 Query Match 55.3%; Score 42; DB 1; Length 386;  
 Best Local Similarity 54.5%; Pred. No. 9.1;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 VSYVHNGAKP 15  
 DB 78 IRYIHANGASF 88  
 RESULT 5  
 ID Y441\_CHLPN STANDARD; PRT; 316 AA.  
 AC Q928A2; Q9JRM4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein CPN0441/CP0312/CPJ0441.  
 GN CPN0441 OR CP0312 OR CPJ0441.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).

CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Greenwood J., Davis R.W., Stephens R.S.,  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uitterlind T., Berry K., Baes S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.,  
 RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia  
 RT pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.,  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0441/CT007/TC0275  
 CC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001627; AAD18585.1; -  
 DR EMBL; AE002194; AAF73653.1; -  
 DR EMBL; AP002546; BAA98649.1; -  
 DR TIGR; CP0312; -  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 243 243 Y -> H (IN REF. 1).  
 SQ SEQUENCE 316 AA; 35415 MW; C821CFB53B014B41 CRC64;  
 Query Match 53.9%; Score 41; DB 1; Length 316;  
 Best Local Similarity 61.5%; Pred. No. 11;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 AEVSYYHNGAKP 15  
 DB 44 AOVYLVKVNDAKF 56  
 RESULT 6  
 ID PHYL\_CERPU STANDARD; PRT; 1307 AA.  
 AC P25848; P93100;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Light-sensor Protein Kinase [includes: Phytochrome; Protein kinase  
 DE (EC 2.7.1.-)].  
 GN PHYL OR PHV.  
 OS Ceratodon purpureus (Moss).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 OC Bryopsida; Dicranidae; Dicranales; Dillitaceae; Ceratodon.  
 OX NCBI\_TaxID=3225;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9309252; PubMed=1463836;  
 RA Thummler F., Dufner M., Kreisel P., Dittich P.  
 RT "Molecular cloning of a novel phytochrome gene of the moss *Ceratodon purpureus* which encodes a putative light-regulated protein kinase."  
 RL Plant Mol. Biol. 20:1003-1017(1992).  
 RN [2]  
 RP SEQUENCE OF 49-538 FROM N.A.  
 RX MEDLINE=91085543; PubMed=2261981;  
 RA Thummler F., Beetz A., Ruediger W.  
 RT "Phytochrome in lower plants. Detection and partial sequence of a phytochrome gene in the moss *Ceratodon purpureus* using the polymerase chain reaction."  
 RL FEBS Lett. 275:125-129(1990).  
 RN [3]  
 RP REVISIONS TO C-TERMINUS.  
 RA STRAIN-WT3;  
 RA Pasenteis K., Paulo N., Dittich P., Algarra P., Thummler F., Dufner M., Kreisel P.  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-BIPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN, PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: LOCATED IN A FIXED POSITION CLOSE TO THE PLASMA MEMBRANE.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U87632; AAB4762.1; -;  
 DR EMBL; X17084; CAA34936.1; ALT\_SEQ.  
 DR PIR; S12966; S12966.  
 DR HSR; S27396; S27396.  
 DR HSP; P08631; IAD5.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR003018; GAF.  
 DR InterPro; IPR000700; PAS-associ\_C.  
 DR InterPro; IPR000014; PAS domain.  
 DR InterPro; IPR001294; Phytochrome.  
 DR InterPro; IPR004040; STY\_Pkinase.  
 DR InterPro; IPR002290; Ser\_Thr\_Pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00360; Phytochrome; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR Pfam; PF01590; GAF; 1.  
 DR PRINTS; PR01033; PHYTOCHROME.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00065; GAF; 1.  
 DR SMART; SM00091; PAS; 1.  
 DR SMART; SM00221; STYK; 1.  
 DR TIGRfam; TIGR00229; sensory\_box; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE\_ST; 1.  
 DR PROSITE; PS00111; PROTEIN KINASE\_DOM; 1.  
 DR PROSITE; PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE; PS00046; PHYTOCHROME\_2; 1.  
 DR PROSITE; PS00112; PAS; 1.  
 DR PROSITE; PS00113; PAC; 1.  
 DR Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 609 680 PAS.  
 FT DOMAIN 683 739 PAC.  
 FT DOMAIN 779 1003 HINGE.  
 FT DOMAIN 1004 1307 PROTEIN KINASE.  
 FT BINDING 320 320 CHROMOPHORE (BY SIMILARITY).  
 FT NP\_BIND 1010 1018 ATP (BY SIMILARITY).  
 FT BINDING 1031 1031 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1127 1127 BY SIMILARITY.  
 SQ SEQUENCE 1307 AA; 145843 MW; E5E77A9FEF301A5C CRC64;  
 Query Match 53.9%; Score 41; DB 1; Length 1307;  
 Best Local Similarity 46.7%; Pred. No. 41;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 SRAEVSYYVNGAKE 15  
 Db 1141 TKSIEGYVHVXADF 1155  
 ID CYB\_L0LBL STANDARD; PRT; 379 AA.  
 AC 047477;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome B.  
 GN COB OR CYTB.  
 OS Lolligo bleekeri (Bleeker's squid).  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;  
 OC Myopsida; Loliginidae; Loligo.  
 OX NCBI\_Taxid=6617;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tomita K., Ueda T., Watanabe K.  
 RT "Completing of squid (*Lolligo bleekeri*) mitochondrial genome sequencing."  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS.  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN.  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B, CYTOCHROME C1 AND THE RIESKE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB009838; BAA2406.1; -;  
 DR EMBL; AB029616; BAB03647.1; -;  
 DR InterPro; IPR000179; Cyt\_b\_b6.  
 DR Pfam; PF00032; cytochrome\_b\_c1; 1.  
 DR Pfam; PF00033; cytochrome\_b\_n; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 DR PROSITE; PS00193; CYTOCHROME\_B\_CO; 1.

KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;  
 Heme.  
 FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).  
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).  
 FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).  
 FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).  
 SQ SEQUENCE 379 AA; 43399 MW; DFFFF1A057B33CA CRC64;

Query Match 52.6%; Score 40; DB 1; Length 379;  
 Best Local Similarity 66.7%; Pred. No. 19;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 YHVNGAKF 15  
 DB 82 YIHANGASF 90

RESULT 8  
 CYB\_DICDI STANDARD; PRT; 389 AA.  
 AC Q37311;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome B.  
 GN COB OR CYTB.  
 OS Dictyostelium discoideum (slime mold).  
 OG Mitochondrion.  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_Taxid=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX3;  
 RX MEDLINE=95254668; PubMed=7736610;  
 RA Angata K., Kuroe K., Yanaigawa K., Tanaka Y.;  
 RT "Codon usage, genetic code and phylogeny of Dictyostelium discoideum  
 mitochondrial DNA as deduced from a 7.3-kb region.";  
 RL Curr. Genet. 27:249-256(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX3;  
 RX MEDLINE=20279206; PubMed=10821186;  
 RA Ogawa S., Yoshino R., Angata K., Iwamoto M., Pi M., Kuroe K.,  
 RA Matsuo K., Morio T., Urushihara H., Yanaigawa K., Tanaka Y.;  
 RT "The mitochondrial DNA of Dictyostelium discoideum: complete sequence,  
 RT gene content and genome organization.";  
 RL Mol. Gen. Genet. 263:514-519(2000).  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS.  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN.  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D14666; BAA03933.1;  
 DR EMBL; AB000109; BAA78061.1;  
 DR DCCYB: DD07272; cob.  
 DR InterPro: IPR000179; Cyt\_b\_b6.  
 DR Pfam: PF00032; cytochrome\_b\_c1.  
 DR Pfam: PF00033; cytochrome\_b\_n\_1.  
 DR PROSITE: PS00192; CYTOCHROME\_B\_HEME; 1.  
 DR PROSITE: PS00193; CYTOCHROME\_B\_OO; 1.

KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;  
 Heme.  
 FT METAL 82 82 IRON 1 (HEME B562 AXIAL LIGAND).  
 FT METAL 96 96 IRON 2 (HEME B566 AXIAL LIGAND).  
 FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).  
 FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).  
 SQ SEQUENCE 389 AA; 44483 MW; PFF9E7092AAEBB1 CRC64;

Query Match 52.6%; Score 40; DB 1; Length 389;  
 Best Local Similarity 66.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 YHVNGAKF 15  
 DB 80 YIHANGASF 88

RESULT 9  
 YTFM\_HAEIN STANDARD; PRT; 578 AA.  
 AC P44038;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein HI0698 precursor.  
 GN HI0698.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;  
 OC Haemophilus.  
 OX NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KM20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerecavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sadek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 RA Gray C., Fountoulakis M.;  
 RT "Two-dimensional map of the proteome of Haemophilus influenzae";  
 RL Electrophoresis 21:411-429(2000).  
 CC -1- SIMILARITY: STRONG, TO E. COLI YTFM.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U32752; AAC22357.1;  
 DR TIGR; H10698;  
 DR InterPro: IPR000184; Bac\_surfing\_D15.  
 DR Pfam: PF01103; Bac\_surface\_Ag; 1.  
 DR Signal; Complete proteome.  
 DR SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 578 PROTEIN HI0698.  
 SQ SEQUENCE 578 AA; 65012 MW; 34F9A189C505876 CRC64;

Query Match 52.6%; Score 40; DB 1; Length 578;

Best Local Similarity 50.0%; Pred. No. 28;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRAEVSYYHNGAK 12  
DB 447 ABAEIGYHRTKG 458

## RESULT 10

PLR\_ACTCH STANDARD; PRT; 467 AA.  
AC P35336;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Polyalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase).  
OS Actinidia chinensis (Kiwi) (Yangtze).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; Ericales; Actinidiaceae; Actinidia.  
OX NCBI\_TaxID=3625;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Deliciosa;  
RX MEDLINE=94302157; PubMed=8029342;  
RA Atkinson R.G., Gardner R.C.;  
RT "A polyalacturonase gene from kiwifruit (Actinidia deliciosa).";  
RL Plant Physiol. 103:669-670(1993).  
CC -1- FUNCTION: ACTS IN CONCERT WITH THE PECTINESTERASE, IN THE RIPENING  
PROCESS. IS INVOLVED IN CELL WALL METABOLISM, SPECIFICALLY IN  
POLYURONIDE DEGRADATION.  
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-  
galactosiduronic linkages in pectate and other galacturonans.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DEVELOPMENTAL STAGE: IN RIPENING FRUIT.  
CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
(POLYGLACTURONASES).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L12019; AAC14453.1; -  
DR InterPro; IPR000743; GH28.  
DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
DR PROSITE; PS00502; POLYGLACTURONASE; 1.  
KW Hydrolyase; Glycosidase; Cell wall; Signal; Fruit ripening;  
KW Glycoprotein.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 467 POLYGLACTURONASE.  
FT ACT SITE 306 306 PROBABLE.  
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 467 AA; 50776 MW; 5A9A61483C028B7A CRC64;  
Query Match 51.3%; Score 39; DB 1; Length 467;  
Best Local Similarity 71.4%; Pred. No. 34;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRAEVSYYHNGAK 14  
DB 318 SEAHVSQVYVNGAK 331

RESULT 11  
CYB\_SORDA STANDARD; PRT; 134 AA.  
AC O21411; O21410;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Cytochrome B (Fragment).  
GN MTCYB OR COB OR CYTB.  
OS Sorex daphaenodon (Large-toothed shrew).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Insectivora; Soricidae; Soricinae; Sorex.  
OX NCBI\_TaxID=62272;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=#696, and 95/8/19-1; TISSUE=Hindfoot;  
RA Oudachi S., Masuda R., Abe H., Adachi J., Dokuchaev N.E.,  
RA Hukusai V., Yoshida M.C.;  
RT "Molecular phylogeny from nucleotide sequences of the mitochondrial  
cytochrome b gene and evolutionary history of Eurasian soricine shrews  
(Mammalia, Insectivora).";  
RL Zool. Sci. 14:527-532(1997).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
COUPLED TO ATP SYNTHESIS  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
BOUND TO THE PROTEIN.  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CYTOCHROME C1 AND THE RISKE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D85356; BAA21349.1; -  
DR EMBL; D85342; BAA21335.1; -  
DR InterPro; IPR00179; Cyt b b6.  
DR Pfam; PF00033; cytochrome b N; 1.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00193; CYTOCHROME\_B\_CO; PARTIAL.  
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;  
KW Heme.  
FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).  
FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).  
FT VARIANT 82 84 LHV -> FHA (IN 95/8/19-1).  
FT NON TER 134 134  
SQ SEQUENCE 134 AA; 15218 MW; 3D53D624E81B7F7C CRC64;

Query Match 50.0%; Score 38; DB 1; Length 134;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 VSYVHNGA 13  
DB 79 IRYLVHNGA 87

RESULT 12  
Y113\_AQUAE STANDARD; PRT; 348 AA.  
AC O66516;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AQ\_113 precursor.  
GN AQ\_113.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-VF5;  
RA MEDLINE=96196666; PubMed=9537320;  
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus".  
RL Nature 392:353-358 (1998).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000674; AAC06480.1; -  
KW Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 348 HYPOTHETICAL PROTEIN AQ\_113.  
SQ SEQUENCE 348 AA; 38760 MW; BDDC09E0013AF152 CRC64;  
Query Match 50.0%; Score 38; DB 1; Length 348;  
Best Local Similarity 57.1%; Pred. No. 38;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 2 RAEVSYVHNGAKF 15  
DB 316 RAEVSYVNTDNKVF 329  
RESULT 13  
CYB\_SACDO STANDARD; PRT; 385 AA.  
ID CYB\_SACDO  
AC Q35819;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Cytochrome B.  
GN COB OR CYTB.  
OS Saccharomyces douglasii (Yeast).  
OC Mitochondrion.  
OC Baktaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxId=46617;  
RN RN  
RP SEQUENCE FROM N.A.  
RA MEDLINE=91218158; PubMed=1708831;  
RA Tian G.L., Michel F., Macadre C., Slonimski P.P., Lazowska J.,  
RT "Incipient mitochondrial evolution in yeasts. II. The complete  
RT sequence of the gene coding for cytochrome b in Saccharomyces  
RT douglasii reveals the presence of both new and conserved introns and  
RT discloses major differences in the fixation of mutations in  
RT evolution."  
RL J. Mol. Biol. 218:747-760 (1991).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC COUPLED TO ATP SYNTHESIS.  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC BOUND TO THE PROTEIN.  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC CYTOCHROME C1 AND THE RISKE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; X59280; CAA41971.1; -  
DR InterPro; IPR000179; Cyt\_b\_b6.  
DR Pfam; PF00032; cytochrome\_b\_c1.  
DR Pfam; PF00033; cytochrome\_b\_n; 1.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00193; CYTOCHROME\_B\_OO; 1.  
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;  
Heme.  
FT METAL 82 82 IRON 1 (HEME B562 AXIAL LIGAND).  
FT METAL 96 96 IRON 2 (HEME B566 AXIAL LIGAND).  
FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).  
FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).  
SQ SEQUENCE 385 AA; 43631 MW; 3541B2C7D77B2DD2 CRC64;  
Query Match 50.0%; Score 38; DB 1; Length 385;  
Best Local Similarity 66.7%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 7 YVHNGAKF 15  
DB 80 YVHNGASF 88  
RESULT 14  
CYB\_YEAST STANDARD; PRT; 385 AA.  
ID CYB\_YEAST  
AC P00163; Q35807; Q36301; Q35802;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cytochrome B.  
GN COB OR CYTB.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxId=4932;  
RN RN  
RP SEQUENCE FROM N.A.  
RA MEDLINE=D273-10B/A21;  
RA MEDLINE=81046788; PubMed=6253454;  
RA Nobrega F.G., Tzagoloff A.,  
RT "Assembly of the mitochondrial membrane system. DNA sequence and  
RT organization of the cytochrome b gene in Saccharomyces cerevisiae  
RT D273-10B".  
RL J. Biol. Chem. 255:9828-9837 (1980).  
RN RN  
RP SEQUENCE FROM N.A.  
RA MEDLINE=D273-10B/A21;  
RA MEDLINE=85000973; PubMed=6383504;  
RA Bonjardim C.A., Nobrega F.G.,  
RT "Revision of the nucleotide sequence at the last intron of the  
RT mitochondrial apocytochrome b gene in Saccharomyces cerevisiae".  
RL Braz. J. Med. Biol. Res. 17:17-20 (1984).  
RN RN  
RP SEQUENCE FROM N.A.  
RA STRAIN=MR200;  
RA MEDLINE=95355283; PubMed=7737175;  
RA Claos M.G., Pera J., Shu Y., Samatey F.A., Popot J.L., Jacq C.,  
RT "Limitations to in vivo import of hydrophobic proteins into yeast  
RT mitochondria. The case of a cytoplasmically synthesized apocytochrome  
RT b".  
RL Eur. J. Biochem. 228:762-771 (1995).  
RN RN  
RP SEQUENCE OF 20-143 FROM N.A.  
RA STRAIN=777-3A;  
RA MEDLINE=81088336; PubMed=7004642;  
RA Lazowska J., Jacq C., Slonimski P.P.,  
RT "Sequence of introns and flanking exons in wild-type and box3 mutants  
RT of cytochrome b reveals an interlaced splicing protein coded by an  
RT intron".  
RL Cell 22:353-348 (1980).

```

RN [5]
RP SEQUENCE OF 144-169 FROM N.A.
RC STRAIN=777-3A;
RX MEDLINE=82115326; PubMed=7034963;
RA Lazowska J., Jacq C., Slonimski P.P.;
RT "Splice points of the third intron in the yeast mitochondrial
   cytochrome b gene.";
RL Cell 27:12-14(1981).
RN [6]
RP MUTANT W7.
RX MEDLINE=90005972; PubMed=2551731;
RA Brivet-Chevillotte P., di Rago J.-P.;
RT "Electron-transfer restoration by vitamin K3 in a complex
   III-deficient mutant of S. cerevisiae and sequence of the
   corresponding cytochrome b mutation.";
RL FEBS Lett. 255:5-9(1989).
CC -1- FUNCTION: COMPONENT OF THE UBIOQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01476; AAA99924.1; ALT_SEQ.
DR EMBL; X84042; CAA58861.1; -.
DR EMBL; V00686; -; NOT ANNOTATED_CDS.
DR EMBL; J01473; AAA32151.2; -.
DR EMBL; J01472; AAA32151.2; JOINED.
DR EMBL; J01475; AAA32152.2; -.
DR EMBL; J01474; AAA32152.2; JOINED.
DR PIR; A00159; CBBY.
DR SGD; S0007270; COB.
DR InterPro: IPR000179; Cyt b_b6.
DR Pfam; PF00032; cytochrome_b_c1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
KV METAL.
FT METAL. 82 82 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL. 96 96 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL. 183 183 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL. 197 197 IRON 1 (HEME B566 AXIAL LIGAND).
FT METAL. 197 197 IRON 1 (HEME B566 AXIAL LIGAND).
FT VARIANT. 131 131 G -> S (IN MUTANT W7 WHICH IS RESPIRATORY
FT DEFICIENT).
FT CONFLICT. 122 122 T -> I (IN REF. 1 AND 2).
FT CONFLICT. 270 270 D -> V (IN REF. 3).
SQ SEQUENCE 385 AA; 43659 MW; 23C41614B712A79C CRC64;

Query Match 50.0%; Score 38; DB 1; Length 385;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 7 YVHVNKAF 15
DB 80 YLHANGASF 88

```

```

DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B.
GN COB OR CYTB.
OS Hansenula wingei (Yeast).
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxId=4907;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21;
RA Sekito T., Okamoto K., Kitano H., Yoshida K.;
RT "Yeast Hansenula wingei mitochondria genome's complete DNA sequence
   demonstrated unique characteristics.";
RL Nucleic Acids Symp. Ser. 31:233-234(1994).
CC -1- FUNCTION: COMPONENT OF THE UBIOQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D31785; BAA06572.1; -.
DR InterPro: IPR000179; Cyt b_b6.
DR Pfam; PF00032; cytochrome_b_c1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
KV METAL.
FT METAL. 82 82 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL. 96 96 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL. 183 183 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL. 197 197 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 386 AA; 44083 MW; 8B46CD4553C643B8 CRC64;

Query Match 50.0%; Score 38; DB 1; Length 386;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 7 YVHVNKAF 15
DB 80 YLHANGASF 88

```

Search completed: April 20, 2003, 13:07:45  
Job time : 3.92105 secs

RESULT 15  
CYB\_HANNI  
ID CYB\_HANNI  
AC P48877;  
STANDARD; PRT; 386 AA.

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:56:55 ; Search time 12.6711 Seconds

(without alignments)  
243.919 Million cell updates/sec

Title: US-09-142-524d-131

Perfect score: 76

Sequence: 1 SRAEVSIVHNGAKP 15

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	80.3	507	10	Q9FY19
2	45	59.2	155	16	Q8UG30
3	43	56.6	431	10	Q23147
4	43	56.6	131	8	Q29131
5	42	55.3	380	8	Q36562
6	42	55.3	497	8	Q36562
7	42	55.3	497	8	Q36562
8	42	55.3	1195	11	Q54852
9	42	55.3	1195	11	Q54852
10	42	55.3	1195	11	Q54852
11	41	53.9	198	17	Q97426
12	41	53.9	563	17	Q97426
13	41	53.9	608	17	Q8ZSM4
14	41	53.9	1299	10	P93098
15	40	52.6	82	10	Q8SAB4
16	40	52.6	120	8	O99125

17	40	52.6	256	9	Q38596
18	40	52.6	293	16	Q98N09
19	40	52.6	313	8	Q48005
20	40	52.6	366	8	Q95005
21	40	52.6	374	8	Q92Y19
22	40	52.6	379	8	Q8WEK9
23	40	52.6	392	8	Q79719
24	40	52.6	395	8	Q95083
25	39	51.3	119	8	Q47889
26	39	51.3	151	16	Q92EP3
27	39	51.3	177	10	Q9M6S5
28	39	51.3	210	3	Q8X138
29	39	51.3	234	16	Q8R9D2
30	39	51.3	372	16	Q92NE5
31	39	51.3	382	8	Q37395
32	39	51.3	383	8	Q9T253
33	39	51.3	386	3	P81381
34	39	51.3	428	2	Q9Z4N5
35	39	51.3	463	10	Q9M6S2
36	39	51.3	475	5	Q9W1Q1
37	39	51.3	539	10	Q9LNE3
38	39	51.3	549	10	Q9LNE3
39	39	51.3	572	10	Q9LVN8
40	39	51.3	576	16	Q8YVH8
41	39	51.3	711	10	Q9SPM1
42	39	51.3	711	10	Q9SPM1
43	39	51.3	4930	16	Q31783
44	38	50.0	63	10	Q9F2N3
45	38	50.0	89	5	Q20941
			97	16	Q31829

## ALIGNMENTS

## RESULT 1

ID Q9FY19 PRELIMINARY; PRT; 507 AA.

AC Q9FY19; 01-MAR-2001 (T-EMBLrel. 16, Created)

DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)

DT 01-UTN-2002 (T-EMBLrel. 21, Last annotation update)

DE Pollen major allergen 2 protein precursor.

GN JNA2.

OS Juniperus ashei (Ozark white cedar).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.

OX NCBI\_TaxId=13101;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE=MALE POLLEN;

RX MEDLINE=20403896; PubMed=10944464;

RA Yokoyama M., Miyahara M., Shimizu K., Kino K., Tanuo H.;

RT "Purification, Identification and cDNA cloning of Jun a 2, the second

KL major allergen of mountain cedar pollen.";

CC Biochem. Biophys. Res. Commun. 275:195-202(2000).

CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES

CC (POLYGALACTURONASES)

CC EMBL; AJ404653; CAC05582.1; -

CC HSSP; P26509; 1BHR.

DR InterPro; IPR000743; GH28.

DR InterPro; IPR000408; Reg\_cnr\_condens.

DR Pfam; PF00295; Glyco\_hydro\_28; 1.

DR PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN\_1.

DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.

KW Cell wall; Glycosidase; Hydrolase; Signal.

FT SIGNAL 1 54 POTENTIAL.

SQ SEQUENCE 507 AA; 55730 MW; 2B2E0A5E5958FE5A CRC64;

QY 1 SRAEVSIVHNGAKP 15

Query Match 80.3%; Score 61; DB 10; Length 507;

Best Local Similarity 73.3%; Pred. No. 0.0051;

Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 291 SRSEVSVHLDGAKF 305

## RESULT 2

ID Q8UG30 PRELIMINARY; PRT; 155 AA.

AC Q8UG30;  
DT 01-JUN-2002 (TREMblrel. 21, Created)  
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE Hypothetical protein Atu1212.  
GN Atu1212 OR AGR\_C\_2237.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970);  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;

SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavin T., Levy R., Li M.-J., McClelland B., Palmeri A., Gordon D.,  
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Krespan W., Perry M.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao Y., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Neeser B.W.,  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58."  
RT C58."  
RL Science 294:2317-2323 (2001).  
RN [2]

SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194; Miller N., Blanchard M.,  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Quirillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Hountel K., Gordon J., Vaudin M., Tarchnouk O., Bpp A., Liu F.,  
RA William C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58."  
RT Science 294:2323-2328 (2001).  
RL EMBL; AE008084; AAL42223.1; -;  
DR EMBL; AE008084; AAL42223.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 155 AA; 16849 MW; C04F0F4540F0E67 CRC64;

Query Match 59.2%; Score 45; DB 16; Length 155;  
Best Local Similarity 61.5%; Pred. No. 1.5;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 RAESVYVHNGAK 14  
Db 132 RADVAVHVSAR 144

## RESULT 3

ID 023147 PRELIMINARY; PRT; 431 AA.

AC 023147;  
DT 01-JUN-1998 (TREMblrel. 05, Created)  
DT 01-JUN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE Endo-polygalacturonase.  
GN PGAL OR ADPGI OR T8H10.110.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC SpERMotophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, COLOMBIA;  
RA Sander L., Child R., Ulvskov P., Albrechtsen M., Joergensen B.,  
RA Borkhardt B.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=CV, LANDSBERG ERRECTA;  
RA Jenkins E.S., Roberts J.A.;  
RT "Deniscence-related expression of an Arabidopsis thaliana gene encoding a polygalacturonase in transgenic plants of Brassica napus."  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RA Benes V., Reichmann S., Borkova D., Anseorge W., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X., Queller F., Salanoubat M.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]

RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES (POLYGALACTURONASES).  
DR EMBL; AJ002532; CAA05525.1; -;  
DR EMBL; AF037367; AAC98923.1; -;  
DR EMBL; AL133248; CAB6108.1; -;  
DR InterPro; IPR000743; GH28  
DR InterPro; IPR004048; Reg\_chir\_condens.  
DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
DR PROSITE; PS00626; RCCL 2; UNKNOWN 1.  
KW Cell wall; Glycosidase; Hydrolase; Signal.

KW SEQUENCE 431 AA; 46572 MW; 0F0F30DF45804FE7 CRC64;

Query Match 56.6%; Score 43; DB 10; Length 431;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRAVSVHNGAKP 15  
Db 299 SKAVSVHNGAKF 313

## RESULT 4

ID 029131 PRELIMINARY; PRT; 456 AA.

AC 029131;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Hypothetical protein AF1134.  
GN AF1134.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=9604943; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson U.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kitznes S., Reich C.I., McNeil L.K., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Goessens J.D., Weiman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-



RT reducing archaeon *Archaeoglobus fulgidus*.  
 RL Nature 390:364-370(1997).  
 DR EMBL; AE001026; MAB90123.1; -  
 DR TIGR; AF1134; -  
 DR InterPro; IPR000379; Ser esters site.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 456 AA, 51748 MW, 37CB62B58CC9357 CRC64;

Query Match 56.6%; Score 43; DB 17; Length 456;  
 Best Local Similarity 72.7%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 VSYVHNGAKF 15  
 |||:||||  
 Db 178 VSYVHNGAKF 188

## RESULT 5

ID 036562 PRELIMINARY; PRT; 131 AA.  
 AC 036562;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, last annotation update)  
 DE Apocytochrome B (Fragment).  
 GN CYTB  
 OS *Williopsis mrakii* (Yeast) (Hansenula mrakii).  
 OC Mitochondrion.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Williopsis.  
 NCBI\_TaxID=4963;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS1707;  
 RX MEDLINE=93204976; PubMed=8455612;  
 RA Fukuhara H., Sor F., Drissi R., Dinovel N., Miyakawa I., Rousset S.,  
 RA Viola A.M.;  
 RT "Linear mitochondrial DNAs of yeast. I. Occurrence and general  
 RT features";  
 RL Mol. Cell. Biol. 13:2309-2314(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS1707;  
 RX MEDLINE=94287716; PubMed=8017108;  
 RA Drissi R., Sor F., Fukuhara H.;  
 RT "Genes of the linear mitochondrial DNA of *Williopsis mrakii*: Coding  
 RT sequences for a maturase-like protein, a ribosomal protein VAR1  
 RT homologue, cytochrome oxidase subunit 2 and methyl tRNA.";  
 RL Yeast 10:391-398(1994).  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL; X66594; CA47157.1; -  
 DR InterPro; IPR000179; Cyt\_b\_b6.  
 DR Pfam; PF00033; cytochrome\_b\_N; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 KW Electron transport; Heme; Mitochondrion; Respiratory chain;  
 KW Transmembrane.  
 FT NON\_TER 131  
 SQ SEQUENCE 131 AA, 14799 MW, 835A67A6B8B23265 CRC64;

Query Match 55.3%; Score 42; DB 8; Length 131;  
 Best Local Similarity 54.5%; Pred. No. 4.5;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 VSYVHNGAKF 15  
 :|||

Db 78 IRYIHANGASF 88

## RESULT 6

ID 047792 PRELIMINARY; PRT; 380 AA.  
 AC 047792;  
 DT 01-JUN-1998 (TRENBLREL. 06, Created)  
 DT 01-JUN-1998 (TRENBLREL. 06, last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, last annotation update)  
 DE Cytochrome b (Fragment).  
 GN CYTB  
 OS *Xenocypis yumanensis*.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Xenocypis.  
 NCBI\_TaxID=70549;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE;  
 RX MEDLINE=21097357; PubMed=11161753;  
 RA Xiao W., Zhang Y., Liu H.;  
 RT "Molecular Systematics of *Xenocypis* (Teleostei: Cyprinidae):  
 RT Taxonomy, Biogeography, and Coevolution of a Special Group Restricted  
 RT in East Asia.";  
 RL Mol. Phylogenet. Evol. 18:163-173(2001).  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 DR EMBL; AF036208; AAC15610.1; -  
 DR InterPro; IPR000179; Cyt\_b\_b6.  
 DR Pfam; PF00033; cytochrome\_b\_N; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 DR PROSITE; PS00193; CYTOCHROME\_B\_OO; UNKNOWN 1.  
 KW Electron transport; Heme; Mitochondrion; Respiratory chain;  
 KW Transmembrane.  
 FT NON\_TER 380  
 SQ SEQUENCE 380 AA, 42988 MW, E5CDD7E9E93D42CF CRC64;

Query Match 55.3%; Score 42; DB 8; Length 380;  
 Best Local Similarity 54.5%; Pred. No. 15;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 VSYVHNGAKF 15  
 :|||

Db 79 IRYIHANGASF 89

## RESULT 7

ID 09684 PRELIMINARY; PRT; 497 AA.  
 AC 09684;  
 DT 01-MAR-2001 (TRENBLREL. 16, Created)  
 DT 01-MAR-2001 (TRENBLREL. 16, last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, last annotation update)  
 DE Apocytochrome b (EC 1.10.2.2).  
 GN CYTB OR COB  
 OS *Naegleria gruberi*.  
 OC Mitochondrion.  
 OC Eukaryota; Eukaryotae; Schizozoa; Vahlkampfiidae; Naegleria.

OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.  
 NCBI\_TaxID=5762;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Burger G., Lang B.F., Neraud T.A., Gray M.W.;  
 RT "The mitochondrial genome of the supposedly primitive protist,

```

RT Naegleria gruberi." to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY). WHICH ARE NOT COVALENTLY
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B563) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF288092; AAG17790.1; -.
DR InterPro; IPR001179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c1; 1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_CO; UNKNOWN; 1.
DR Electon transport; Heme; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
SQ SEQUENCE 497 AA; 58323 MW; 7A77A3FEB8F690E2 CRC64;

Query Match 55.3%; Score 42; DB 8; Length 497;
Best Local Similarity 54.5%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 VSYVHNGAKF 15
DB 82 IRYIHNGASF 92

RESULT 8
ID 054852 PRELIMINARY; PRT; 1195 AA.
AC 054852;
DT 01-JUN-1998 (TREMBLrel. 06, Created).
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Potassium channel.
GN ERG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98054206; PubMed=9390998;
RA Shi W., Wymore R.S., Wang H.S., Pan Z., Cohen I.S., McKinnon D.,
RA Dixon J.E.;
RT "Identification of two nervous system-specific members of the erg
RT potassium channel gene family."
RL J. Neurosci. 17:9423-9432(1997).
DR EMBL; AF016191; AAB94741.1; -.
DR InterPro; IPR000595; CNMP_binding.
DR InterPro; IPR003967; Erg_channel.
DR InterPro; IPR001622; K+channel_pore.
DR InterPro; IPR000636; M+channel_nlg.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00027; CNMP_binding; 1.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF00785; PAC; 1.
DR PRINTS; PR01470; ERGCHANNEL.
DR SMART; SM00100; CNMP; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ SEQUENCE 1195 AA; 134900 MW; BA24C54BB86C59B7 CRC64;

Query Match 55.3%; Score 42; DB 11; Length 1195;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 RA6SYVHNGAKF 15

```

```

DB 93 KVEVTVYHNGSTF 106

RESULT 9
ID 09ER47 PRELIMINARY; PRT; 1195 AA.
AC 09ER47;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Erg3 protein.
GN ERG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Arcangelii A.;
RT "Erg genes expression during development of mouse embryos."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; A291608; CAC14797.1; -.
DR InterPro; IPR000595; CNMP_binding.
DR InterPro; IPR001622; K+channel_pore.
DR InterPro; IPR000636; M+channel_nlg.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR00027; CNMP_binding; 1.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF00785; PAC; 1.
DR SMART; SM00100; CNMP; 1.
DR SMART; SM00086; PAC; 1.
DR PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ SEQUENCE 1195 AA; 135026 MW; 8DCCB9BA0580FFC2 CRC64;

Query Match 55.3%; Score 42; DB 11; Length 1195;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 RA6SYVHNGAKF 15
DB 93 KVEVTVYHNGSTF 106

RESULT 10
ID 09NS40 PRELIMINARY; PRT; 1196 AA.
AC 09NS40;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Potassium channel subunit.
GN HERG-3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Ganetzky B., Titus S.A.;
RT "Polynucleotides encoding herg-3 potassium channel."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF032897; AAD01946.1; -.
DR InterPro; IPR000595; CNMP_binding.
DR InterPro; IPR003967; Erg_channel.
DR InterPro; IPR001622; K+channel_pore.
DR InterPro; IPR000636; M+channel_nlg.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00027; CNMP_binding; 1.
DR Pfam; PF00520; Ion_trans; 1.

```

DR Pfam: PF00785; PAC; 1.  
 DR PRINTS: PR01470; ERCHANNEL.  
 DR SMART: SM00100; CNMP; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAC; 1.  
 DR PROSITE: PS50042; CNMP BINDING 3; 1.  
 SO SEQUENCE 1196 AA; 135012 MW; 7CE10C1A176DC4FE CRC64;

Query Match 55.3%; Score 42; DB 4; Length 1196;  
 Best Local Similarity 50.0%; Pred. No. 54;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 RAESYVHNGAKF 15  
 DB 93 KVEVTVYHKGSTF 106

RESULT 11  
 097VZ6 PRELIMINARY; PRT; 198 AA.

AC 097VZ6; PRELIMINARY; PRT; 198 AA.  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Endonuclease V (nfi) (EC 3.1.25.1).  
 GN NFI OR SSO2454  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 NCBI\_Taxid=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aways M.J., Chan-Weiner C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,  
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL: AE006844; AKA42594.1;  
 KW Hydrolyase; Complete proteome.  
 SO SEQUENCE 198 AA; 22051 MW; DE752959A649398B CRC64;

Query Match 53.9%; Score 41; DB 17; Length 198;  
 Best Local Similarity 42.9%; Pred. No. 11;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRAEVSIVHNGAK 14  
 DB 139 NESRITVYVINGEK 152

RESULT 12  
 09HJZ3 PRELIMINARY; PRT; 563 AA.

AC 09HJZ3; PRELIMINARY; PRT; 563 AA.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Hypothetical protein Ta0818.  
 GN TA0818.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 NCBI\_Taxid=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479972; PubMed=11029001;

RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Fishman D., Stocker S., Lups A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 RT acidophilum."  
 RL Nature 407:508-513 (2000).  
 DR EMBL: AL445065; CAC11947.1;  
 DR InterPro: IPR004165; CoA\_trans.  
 DR Pfam: PF01144; CoA\_trans; 1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 563 AA; 62138 MW; A144326E0D7129F5 CRC64;

Query Match 53.9%; Score 41; DB 17; Length 563;  
 Best Local Similarity 50.0%; Pred. No. 36;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 RAESYVHNGAKF 15  
 DB 165 RANVSLVHIDGEEY 178

RESULT 13  
 08ZSM4 PRELIMINARY; PRT; 608 AA.

AC 08ZSM4; PRELIMINARY; PRT; 608 AA.  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Protease IV, conjectural.  
 GN PAB3679.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 NCBI\_Taxid=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX PubMed=11792869;  
 RA Filt-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL: AB009945; AAL65089.1;  
 DR InterPro: IPR002142; Peptidase\_U7.  
 DR Pfam: PF01343; Peptidase\_U7; 1.  
 DR ProDom: PD002897; Peptidase\_U7; 1.  
 KW Complete proteome.  
 SO SEQUENCE 608 AA; 66946 MW; F7357A6501ADCE61 CRC64;

Query Match 53.9%; Score 41; DB 17; Length 608;  
 Best Local Similarity 61.5%; Pred. No. 39;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AEVSIVHNGAKF 15  
 DB 492 ANMQYVHNGSLF 504

RESULT 14  
 P93098 PRELIMINARY; PRT; 1299 AA.

AC P93098; PRELIMINARY; PRT; 1299 AA.  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Ceratodaptor (Fragment).  
 OS Ceratodon purpureus (Moss).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 OC Bryopsida; Dicranidae; Dicranales; Ditrichaceae; Ceratodon.  
 NCBI\_Taxid=3225;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MTJ;

RA Pasentsis K., Paulo N., Dittich P., Algarra P., Thuenmler F.;  
 RT "Characterization and expression of the phytochrome gene family in the  
 RT moss *Ceratodon purpureus* (Hedw.) Brid.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: BELONGS TO THE GNR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.

DR EMBL; Y10901; CAA71838.1; -.  
 DR HSSP; P08631; IAD5.  
 DR InterPro; IPR002106; ALCRNA\_ligase1.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR003018; GAF.  
 DR InterPro; IPR000524; HTH\_Gntr.  
 DR InterPro; IPR000700; PAS-aseoc\_C.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR InterPro; IPR001294; Phytochrome.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF01590; GAF; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR Pfam; PF00360; phytochrome; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00035; HTGNTN.  
 DR PRINTS; PR01033; PHYTOCHROME.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00065; GAF; 1.  
 DR SMART; SM00091; PAS; 1.  
 DR SMART; SM00221; STYK; 1.  
 DR TIGRFAMs; TIGR00229; sensory\_box; 1.  
 DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 DR PROSITE; PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE; PS50046; PHYTOCHROME\_2; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR DNA-binding; Receptor; Serine/threonine-protein kinase;  
 KW Transcription regulation.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1299 AA; 144706 MW; 85BDFC2A63212EFA CRC64;

Query Match 53.9%; Score 41; DB 10; Length 1299;  
 Best Local Similarity 46.7%; Pred. No. 92;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRAVSVYHVGAKF 15  
 DB 1133 TKSEIGYHVXVADF 1147

RESULT 15

ID Q8SAB4 PRELIMINARY; PRT; 82 AA.  
 AC Q8SAB4;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Polyalacturonase (EC 3.2.1.15) (Fragment).  
 OS Capsicum frutescens (chili pepper).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.  
 OX NCBI\_TaxID=4073;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rao G.U., Paran I.;  
 RT "Polygalacturonase: A candidate gene for the soft flesh and deciduous  
 RT fruit mutation in Capsicum."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF465938; AAL76254.1; -.  
 KW Hydrolyase; Glycosidase.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 82 AA; 8384 MW; F2EC39B2C4CEB92 CRC64;

Query Match 52.6%; Score 40; DB 10; Length 82;  
 Best Local Similarity 60.0%; Pred. No. 6.3; 4; Indels 0; Gaps 0;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 SRAVSVYHVGAKF 15  
 DB 54 SEAHVSDINYGAKF 68

Search completed: April 20, 2003, 13:13:21  
 Job time: 15.6711 secs

GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:57:30 ; Search time 6.07895 Seconds

(without alignments)  
237.215 Million cell updates/sec

Title: US-09-142-524d-152

Perfect score: 67

Sequence: 1 LSDISLKLTSKGIAS 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR.73.\*

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	100.0	514	2	Cry j II protein -
2	67	100.0	514	2	second major aller
3	54	80.6	514	2	polygalacturonase
4	51	76.1	507	2	Jun a 2 protein -
5	45	67.2	251	2	high-affinity zinc
6	44	65.7	242	2	ABC transporter, A
7	44	65.7	242	2	hypothetical prote
8	43	64.2	251	2	probable ABC trans
9	43	64.2	251	2	hypothetical prote
10	43	64.2	251	2	probable ABC trans
11	42	62.7	255	1	membrane-bound iro
12	40	59.7	133	2	conserved hypotet
13	40	59.7	134	2	hypothetical 14.2K
14	40	59.7	253	2	high-affinity zinc
15	40	59.7	239	2	hypothetical prote
16	40	59.7	317	2	hypothetical prote
17	39	58.2	93	2	hypothetical prote
18	39	58.2	292	2	hypothetical prote
19	39	58.2	292	2	probable transcrip
20	39	58.2	335	2	hypothetical prote
21	39	58.2	839	2	myosin heavy chain
22	38	56.7	226	2	ABC transporter, A
23	38	56.7	226	2	ABC transporter, A
24	38	56.7	230	2	ABC transporter, A
25	38	56.7	230	2	ABC transporter, A
26	38	56.7	230	2	ABC transporter, A
27	38	56.7	230	2	ABC transporter, A
28	38	56.7	230	2	ABC transporter, A
29	38	56.7	230	2	ABC transporter, A

## ALIGNMENTS

30	38	56.7	398	2	C84780	hypothetical prote
31	38	56.7	505	2	C90569	hypothetical prote
32	38	56.7	529	2	T42584	legume protein 4
33	38	56.7	350	1	WZBED4	gene 40 protein -
34	38	56.7	548	2	T47548	hypothetical prote
35	37	55.2	142	2	H69745	hypothetical prote
36	37	55.2	165	2	F89835	hypothetical prote
37	37	55.2	231	2	C70000	hypothetical prote
38	37	55.2	232	2	C69822	ABC transporter (A
39	37	55.2	240	2	G71905	glycine betaine/L-
40	37	55.2	240	2	C64609	ABC transporter, A
41	37	55.2	253	2	H89845	hypothetical prote
42	37	55.2	254	2	A95410	hypothetical prote
43	37	55.2	255	2	AC0023	probable ABC trans
44	37	55.2	269	2	D75631	probable taurine t
45	37	55.2	283	2	T44799	iron ABC transport

## RESULT 1

S48730 Cry j II protein - Japanese cedar

C;Species: Cryptomeria japonica (Japanese cedar)

C;Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999

C;Accession: S48730

R;Name: M.; Kurose, M.; Toriue, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.;

FEBS Lett. 353, 124-128, 1994

A;Title: Molecular cloning of the second major allergen, Cry j II, from Japanese ceda

A;Reference number: S48730; MUID:95010777; PMID:7926035

A;Accession: S48730

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-514 <KMW>

A;Cross-References: GB:D37765; NID:9577695; PIDN:BA07021.1; PID:dl007598; PID:957769

Query Match

Best Local Similarity 100.0%; Score 67; DB 2; Length 514;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LSDISLKLTSKGIAS 15

Db 395 LSDISLKLTSKGIAS 409

RESULT 2

JC2498

second major allergen Cry j II precursor - Japanese cedar

C;Species: Cryptomeria japonica (Japanese cedar)

C;Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000

C;Accession: JC2498; A60147

R;Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.

Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994

A;Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese

A;Reference number: JC2498; MUID:94271186; PMID:8002972

A;Accession: JC2498

A;Molecule type: mRNA

A;Cross-References: DDBJ:D29772; NID:9506857; PIDN:BA06172.1; PID:9506858

A;Accession: PC346

A;Molecule type: protein

A;Residues: 52-61 <KO2>

R;Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matunashi, T.

Allergy 45, 309-312, 1990

A;Title: Identification of the second major allergen of Japanese cedar pollen.

A;Reference number: A60147; MUID:90342988; PMID:2382797

A;Accession: A60147

A;Molecule type: protein

A;Residues: 55-64 <SAK>

C;Keywords: glycoprotein; pollen

F;1-54/Domain: signal sequence

F;55-460/Product: second major allergen Cry j #status predicted <STG>

F;55-460/Product: second major allergen Cry j #status predicted <MAT>

F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 67; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LSDISLKTSGKIAS 15  
DB 395 LSDISLKTSGKIAS 409

## RESULT 3

UC7100

polygalacturonase Cha o 2 - Japanese cypress

C/Species: Chamaecyparis obtusa (Japanese cypress)

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

A/Accession: UC7100; PC7026

R/Mori, T.; Yokoyama, M.; Komiyama, N.; Okano, M.; Kino, K.

Biochem. Biophys. Res. Commun. 263, 166-171, 1999

A/Title: Purification, identification, and cDNA cloning of Cha o 2, the second major all

A/Reference number: UC7100; MUID:19417540; PMID:10486272

A/Molecule type: mRNA

A/Residues: 1-514 <MOR>

A/Accession: PC7026

A/Molecule type: protein

A/Residues: 51-62 <MO2>

Query Match 80.6%; Score 54; DB 2; Length 514;  
Best Local Similarity 80.0%; Pred. No. 0.044;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LSDISLKTSGKIAS 15  
DB 395 LSDISLKTSGKIAS 409

## RESULT 4

UC7366

Jun a 2 protein - mountain cedar

C/Species: Juniperus ashei (mountain cedar)

C/Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

A/Accession: UC7366; PC7093

R/Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.

Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A/Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A/Reference number: UC7366

A/Accession: UC7366

A/Molecule type: mRNA

A/Residues: 1-507 <YOK>

A/Cross-references: GB:AJ404653

A/Accession: PC7093

A/Molecule type: protein

A/Residues: 55-63 <YOK>

C/Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

o the polygalacturonase family.

C/Keywords: glycoprotein; pollen

Query Match 76.1%; Score 51; DB 2; Length 507;  
Best Local Similarity 66.7%; Pred. No. 0.16;  
Matches 10; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 LSDISLKTSGKIAS 15  
DB 396 LSDISLKTSGKIAS 410

## RESULT 5

AH0742

high-affinity zinc uptake system ATP-binding protein [imported] - Salmonella enterica s

C/Species: Salmonella enterica subsp. enterica serovar Typhi

A/Note: This species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 17-May-2002

C/Accession: AH0742

R/Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church

Ln, T.; Comerford, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser

A/Reference number: AB0502; PMID:11677608

A/Accession: AH0742

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-251 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD05643.1; PID:G16503139; GSPDB:GN00176

C/Genetics:

A/Genes: STY2100

C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 67.2%; Score 45; DB 2; Length 251;

Best Local Similarity 69.2%; Pred. No. 1;

Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 LSDISLKTSGKI 13  
DB 20 LSDVSLSPSKI 32

## RESULT 6

P95256

ABC transporter, ATP-binding protein SP2196 [imported] - Streptococcus pneumoniae (str

C/Species: Streptococcus pneumoniae

C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

A/Accession: P95256

R/Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; He

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf

son, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A/Reference number: A95000; MUID:21357209; PMID:11463916

A/Accession: P95256

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-242 <KOR>

A/Cross-references: GB:AE005672; PIDN:AAK76247.1; PID:G14973708; GSPDB:GN00164; TIGR

A/Experimental source: strain TIGR4

C/Genetics:

A/Genes: SP2196

Query Match 65.7%; Score 44; DB 2; Length 242;  
Best Local Similarity 53.3%; Pred. No. 1.5;  
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 LSDISLKTSGKIAS 15  
DB 19 LEDINLVOTSGEVS 33

## RESULT 7

G98121

hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)

C/Species: Streptococcus pneumoniae

C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

A/Accession: G98121

R/Hockins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A/Reference number: A97872; MUID:21429245; PMID:11544234

A/Accession: G98121

A/Status: preliminary

A/Molecule type: DNA

A;Residues: 1-242 <K12>  
 A;Cross-references: GB:AE007317; PIDN:AA00804.1; PID:g15459706; GSPDB:GN00174  
 C;Genetics:  
 A;Gene: ABC-NBD

Query Match 65.7%; Score 44; DB 2; Length 242;  
 Best Local Similarity 53.3%; Pred. No. 1.5;  
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LSDISLKLTKSGKIAS 15  
 19 LEDIVLQVTSSEVVS 33

RESULT 8  
 B64948  
 Probable ABC transport system ATP-binding protein yebM - Escherichia coli (strain K-12)  
 C;Species: Escherichia coli  
 C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C;Accession: B64948  
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A;Title: The complete genome sequence of Escherichia coli K-12.  
 A;Reference number: A64720; MUID:97426617; PMID:9278503  
 A;Accession: B64948  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-251 <BLAT>  
 A;Cross-references: GB:AE000280; GB:U00096; NID:g178165; PIDN:AACT4928.1; PID:g178165;  
 A;Experimental source: strain K-12, substrain MG1655  
 C;Genetics:  
 A;Gene: yebM  
 C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
 C;Keywords: ATP; nucleotide binding; P-loop; transport protein  
 F;20-197/Domain: ATP-binding cassette homology <ABC>  
 F;37-44/Region: nucleotide-binding motif A (P-loop)

Query Match 64.2%; Score 43; DB 2; Length 251;  
 Best Local Similarity 69.2%; Pred. No. 2.5;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LSDISLKLTKSGKI 13  
 20 LSDVSLKLKPKKI 32

RESULT 9  
 D85798  
 Hypothetical protein yebM (imported) - Escherichia coli (strain O157:H7, substrain EDL93  
 C;Species: Escherichia coli  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
 C;Accession: D85798  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: D85798  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-251 <STO>  
 A;Cross-references: GB:AE005174; NID:g12515914; PIDN:AA656848.1; GSPDB:GN00145; UWGP:Z29  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: yebM  
 C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 64.2%; Score 43; DB 2; Length 251;  
 Best Local Similarity 69.2%; Pred. No. 2.5;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LSDISLKLTKSGKI 13

DB 20 LSDVSLKLKPKKI 32

RESULT 10  
 H90949  
 Probable ABC transport system ATP-binding protein Ecs2568 (imported) - Escherichia co  
 C;Species: Escherichia coli  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 27-Nov-2001  
 C;Accession: H90949  
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
 gasawara, N.; Yasunaga, T.; Kubara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
 A;Reference number: A99629; MUID:2116231; PMID:11258796  
 A;Accession: H90949  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-251 <HAY>  
 A;Cross-references: GB:BA000007; PIDN:BAH35991.1; PID:g13362036; GSPDB:GN00154  
 C;Experimental source: strain O157:H7, substrain RIMD 050952  
 C;Genetics:  
 A;Gene: Ecs2568

Query Match 64.2%; Score 43; DB 2; Length 251;  
 Best Local Similarity 69.2%; Pred. No. 2.5;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LSDISLKLTKSGKI 13  
 20 LSDVSLKLKPKKI 32

RESULT 11  
 QRECM3  
 Membrane-bound iron (III) dicitrate transport protein - Escherichia coli (strain K-12)  
 C;Species: Escherichia coli  
 C;Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 01-Mar-2002  
 C;Accession: J50115; S56512; A65242  
 R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from  
 A;Reference number: S56314; MUID:95534362; PMID:7610040  
 A;Accession: S56512  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-255 <STAS>  
 A;Cross-references: GB:M26397; NID:g145923; PIDN:AAA33765.1; PID:g145928  
 A;Experimental source: strain K12  
 R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from  
 A;Reference number: S56314; MUID:95534362; PMID:7610040  
 A;Accession: S56512  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-255 <BNA>  
 A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97183.1; PID:g537128  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 Science 277, 1453-1462, 1997  
 A;Title: The complete genome sequence of Escherichia coli K-12.  
 A;Reference number: A64720; MUID:97426617; PMID:9278503  
 A;Accession: A65242  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-255 <BLAT>  
 A;Cross-references: GB:AE000499; GB:U00096; NID:g1790732; PIDN:AACT7243.1; PID:g179073  
 A;Experimental source: strain K-12, substrain MG1655  
 C;Genetics:  
 A;Gene: fecC  
 A;Map position: 7 min  
 C;Function:

A:Description: one of five, encoded by the fec operon, constituting a citrate-dependent C:Superfamily: inner membrane protein malK; ATP-binding cassette homology C:Keywords: ATP; iron transport; membrane protein; nucleotide binding; P-loop P:18-214/Domain: ATP-binding cassette homology <ABC> F:35-42/Region: nucleotide-binding motif A (P-loop)

Query Match 62.7%; Score 42; DB 1; Length 255;  
Best Local Similarity 53.3%; Pred. No. 3.9;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGIAS 15  
DB 18 LNDVSLSLPTGKITA 32

## RESULT 12

AI2834  
Conserved hypothetical protein Atu2104 [imported] - Agrobacterium tumefaciens (strain C5

C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AI2834

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Kang, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AI2834

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-133 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAI43095.1; PID:g17740566; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu2104

A:Map position: circular chromosome

Query Match 59.7%; Score 40; DB 2; Length 133;  
Best Local Similarity 40.0%; Pred. No. 4.6;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGIAS 15  
DB 98 MTDLHLKITEGNVAA 112

## RESULT 13

D97612  
Hypothetical 14.2K protein in vacB-aidB intergenic region (o132) [imported] - Agrobacter

C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: D97612

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: D97612

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-133 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87653.1; PID:g15157239; GSPDB:GN00169

A:Gene: AGR\_C 3816

A:Map position: circular chromosome

Query Match 59.7%; Score 40; DB 2; Length 134;  
Best Local Similarity 40.0%; Pred. No. 4.6;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGIAS 15

DB 99 MTDLHLKITEGNVAA 113

## RESULT 14

AD0251  
high-affinity zinc uptake system ATP-binding protein [imported] - Yersinia pestis (str

C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AD0251

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.

deno-Tarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; PMID:21470413; PMID:11586360

A:Accession: AD0251

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-253 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC90872.1; PID:g15980071; GSPDB:GN00175

A:Gene: znuC

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 59.7%; Score 40; DB 2; Length 253;  
Best Local Similarity 69.2%; Pred. No. 9.3;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGI 13  
DB 21 LNDISLRLPGKI 33

## RESULT 15

H90033  
Hypothetical protein SA2132 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: H90033

R:Kuroda, M.; Ohka, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C.; Shiba, T.; Hattori, M.; Ogawara, N.; Hayashi, H.; Hiratake, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: AB9758; PMID:21311952; PMID:11418146

A:Accession: H90033

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-299 <KUR>

A:Cross-references: GB:BA000018; PID:g13702141; PIDN:BAH43433.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA2132

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 59.7%; Score 40; DB 2; Length 299;  
Best Local Similarity 61.5%; Pred. No. 11;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 LSDISLKLTSKGI 13  
DB 18 VNDISLRLSGKM 30

Search completed: April 20, 2003, 13:15:59  
Job time: 8.07895 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524D-100  
Perfect score: 82  
Sequence: 1 NNRIMQFAKLTGFT 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	15	17	AA97887
2	82	100.0	105	18	AAW27370
3	82	100.0	134	18	AAW27371
4	82	100.0	460	16	AA69791
5	82	100.0	514	15	AA53690
6	82	100.0	514	16	AA74333
7	82	100.0	514	16	AA69792
8	82	100.0	514	17	AA93599
9	82	100.0	514	17	AA81586
10	82	100.0	514	20	AA25666

11	82	100.0	514	20	AA25667	Japanese cedar all
12	76	92.7	17	17	AA81581	Cedar pollen aller
13	70	85.4	17	17	AA81588	Cedar pollen aller
14	70	85.4	18	19	AA80346	Sugi allergen prot
15	65	79.3	13	18	AAW12342	Japanese cedar pol
16	61	74.4	12	18	AAW12337	Japanese cedar pol
17	61	74.4	13	18	AAW12536	Japanese cedar pol
18	61	74.4	14	18	AAW12535	Japanese cedar pol
19	58	70.7	20	19	AAW42171	Japanese cedar pol
20	58	70.7	514	17	AAW04346	T-cell epitope pep
21	58	70.7	15	17	AAW42122	Chamaecyparis obtu
22	56	68.3	15	17	AAW97886	Japanese cypress p
23	56	68.3	15	19	AAW57761	Japanese cedar pollen
24	54	65.9	10	17	AAW81575	Residues 76-90 of
25	54	65.9	17	17	AAW81575	Cedar pollen aller
26	52	63.4	453	22	AAW51693	Juniper pollen aller
27	52	63.4	507	22	AAW51691	Uniperus ashei N-
28	50	61.0	11	18	AAW12543	Japanese cedar pol
29	50	61.0	12	18	AAW12544	Japanese cedar pol
30	50	61.0	15	17	AAW97888	Japanese cedar pol
31	47	57.3	436	21	AAW52347	Arabidopsis thalia
32	47	57.3	459	21	AAW52346	Arabidopsis thalia
33	47	57.3	459	23	AAW91108	Herbicideally activ
34	46	56.1	10	18	AAW12545	Japanese cedar pol
35	46	56.1	11	18	AAW12546	Japanese cedar pol
36	46	56.1	12	18	AAW12547	Japanese cedar pol
37	46	56.1	12	22	AAW69107	Cryptomeria japoni
38	46	56.1	93	21	AAW33891	Artificial sequenc
39	46	56.1	93	22	AAW69097	Cedar pollen aller
40	46	56.1	93	22	AAW69097	Cedar pollen aller
41	46	56.1	93	21	AAW33891	Artificial sequenc
42	46	56.1	95	21	AAW33887	Artificial sequenc
43	46	56.1	95	21	AAW33901	Artificial sequenc
44	46	56.1	95	22	AAW69098	Cedar pollen aller
45	46	56.1	96	22	AAW69104	Cedar pollen aller

#### ALIGNMENTS

RESULT 1  
AA97887 standard; peptide; 15 AA.  
AA97887;  
16-AUG-1996 (first entry)  
Japan cedar pollen mature allergen Cry j II amino acids 81-95.  
Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
Sugi pollinosis; diagnosis; treatment.  
Cryptomeria japonica.  
JP08047392-A.  
20-FEB-1996.  
07-NOV-1994; 94JP-0297840.  
26-MAY-1994; 94JP-0134868.  
05-NOV-1993; 93JP-0276773.  
(MEIP) MEIJI MILK PROD CO LTD.  
WPI, 1996-166249/17.  
Japan cedar pollen allergen Cry j II epitope - comprises at least  
part of specified 460 amino acid protein  
Claim 8; Fig 3; 17pp; Japanese.

CC AAR97871-R97960 are overlapping peptides used for the epitope mapping  
CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
CC peptides of it are useful in the diagnosis, prevention and treatment  
CC of Sugi/pollinosis, the allergic reaction to Japan cedar pollen.  
CC Significant regions of the allergen were identified using the  
CC overlapping peptides of the full epitope derived from a Cry j II  
CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200  
CC (R978908) of the full mature 460 amino acid allergen are the most  
CC allergenic of the 90 peptides tested.

XX Sequence 15 AA;

Query Match 100.0%; Score 82; DB 17; Length 15;  
Best Local Similarity 100.0%; Pred. No. 9.7e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNRIMLOFAKLTGFT 15  
DB 1 NNRIMLOFAKLTGFT 15

RESULT 2  
AAM27370  
ID AAM27370 standard; peptide; 105 AA.

XX AAM27370;

XX 24-MAR-1998 (first entry)

XX Multi-epitope peptide used as immunotherapeutic agent #2.

XX Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
XX T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

XX Synthetic.

XX WO9732600-A1.

XX 12-SEP-1997.

XX 10-MAR-1997; 97WO-JP00740.

XX 10-MAR-1996; 96JP-0080702.

XX (MEIP) MEIDI MILK PROD CO LTD.

XX Dairiki K, Iwama A, Kino K, Kume A, Some T;

XX WPI; 1997-470495/43.

XX Peptide immuno:therapeutic agent to treat allergic diseases -  
XX contains multi-epitope peptide containing T cell epitope regions  
XX from different allergens

XX Claim 6; Page 31; 58pp; Japanese.

XX The present sequence represents a multi-epitope peptide which is used as  
XX a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
XX or more different allergens (preferably linked via arginine or lysine  
XX dimers), where the T cell epitope regions have a positivity index  
XX greater than 100 as measured in a patient group responding to the  
XX allergen; have at least 70% reactivity with lymphocytes from patients  
XX responding to the allergen; and are not reactive with immunoglobulin E  
XX (IgE) antibodies from patients responsive to the allergen. The agent can  
XX be used to prevent and treat a wide variety of allergic diseases, e.g. by  
XX desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

XX Sequence 105 AA;

Query Match 100.0%; Score 82; DB 18; Length 105;  
Best Local Similarity 100.0%; Pred. No. 8.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNRIMLOFAKLTGFT 15  
DB 54 NNRIMLOFAKLTGFT 68

RESULT 3  
AAM27371  
ID AAM27371 standard; peptide; 134 AA.

XX AAM27371;

XX 24-MAR-1998 (first entry)

XX Multi-epitope peptide used as immunotherapeutic agent #3.

XX Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
XX T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

XX Synthetic.

XX WO9732600-A1.

XX 12-SEP-1997.

XX 10-MAR-1997; 97WO-JP00740.

XX 10-MAR-1996; 96JP-0080702.

XX (MEIP) MEIDI MILK PROD CO LTD.

XX Dairiki K, Iwama A, Kino K, Kume A, Some T;

XX WPI; 1997-470495/43.

XX Peptide immuno:therapeutic agent to treat allergic diseases -  
XX contains multi-epitope peptide containing T cell epitope regions  
XX from different allergens

XX Claim 6; Page 32; 58pp; Japanese.

XX The present sequence represents a multi-epitope peptide which is used as  
XX a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
XX or more different allergens (preferably linked via arginine or lysine  
XX dimers), where the T cell epitope regions have a positivity index  
XX greater than 100 as measured in a patient group responding to the  
XX allergen; have at least 70% reactivity with lymphocytes from patients  
XX responding to the allergen; and are not reactive with immunoglobulin E  
XX (IgE) antibodies from patients responsive to the allergen. The agent can  
XX be used to prevent and treat a wide variety of allergic diseases, e.g. by  
XX desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

XX Sequence 134 AA;

Query Match 100.0%; Score 82; DB 18; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNRIMLOFAKLTGFT 15  
DB 54 NNRIMLOFAKLTGFT 68

RESULT 4

AAR69791  
ID AAR69791 standard; Protein; 460 AA.

XX AAR69791;

XX 27-SEP-1995 (first entry)

XX Japonicum allergen residues 55-514.

XX Japonicum allergen; residues 55-514; induced histamine release;

KW antiallergic peptide; IGE cross-linking inhibition.  
 XX Japonicum sp.  
 OS  
 XX WO9502412-A.  
 XX  
 XX 26-JAN-1995.  
 PD  
 XX 15-JUL-1994; 94WO-JP01164.  
 XX  
 XX 16-JUL-1993; 93JP-0177008.  
 PR 01-SEP-1993; 93JP-0217725.  
 PR 07-APR-1994; 94JP-0069336.  
 XX  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 XX  
 XX Kino K, Kohno Y, Komiyama N, Sone T;  
 PI  
 XX WPI; 1995-067159/09.  
 DR N-PSDB; AAQ84044.  
 DR  
 XX Peptide antiallergic agent - inhibits cross-linking of allergen  
 PT with IGE antibody  
 XX  
 XX Disclosure; Pages 26-27; 46pp; Japanese.  
 PS  
 CC AAQ84044 encodes AAR69791 Japonicum allergen residues 55-514, from  
 CC which the antiallergic peptides AAR69845-R69809 were derived.  
 CC The peptides ability to inhibit the cross-linking of an allergen,  
 CC to an IGE antibody can be used in the prevention and treatment of  
 CC allergic diseases.  
 CC  
 XX  
 SQ Sequence 460 AA;  
 Query Match 100.0%; Score 82; DB 16; Length 460;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NNRIMLOFAKLTGFT 15  
 DB 81 NNRIMLOFAKLTGFT 95

RESULT 5  
 AAR53690  
 ID AAR53690 standard; Protein; 514 AA.  
 XX  
 AC AAR53690;  
 XX  
 DT 01-FEB-1995 (first entry)  
 XX  
 DE Japanese cedar pollen allergen Cry j II.  
 XX  
 KW Cedar pollinosis; diagnostic.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 XX WO9411512-A.  
 XX  
 XX 26-MAY-1994.  
 PD  
 XX 12-NOV-1993; 93WO-US11000.  
 PF  
 XX 12-NOV-1992; 92US-0975179.  
 PR  
 XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PA  
 XX Brauer A, Kuo M, Pollock J, Yeung S;  
 PI  
 XX WPI; 1994-183513/22.  
 DR N-PSDB; AAQ6048.  
 DR  
 XX Allergenic Cry j II protein and fragments from Japanese cedar

PT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 XX  
 XX Claim 2; Fig 4; 89pp; English.  
 PS  
 XX The sequence is of a Japanese cedar pollen allergen Cry j  
 CC II. The protein and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also AAR53692-6.  
 XX  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 82; DB 15; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NNRIMLOFAKLTGFT 15  
 DB 135 NNRIMLOFAKLTGFT 149

RESULT 6  
 AAR74333  
 ID AAR74333 standard; Protein; 514 AA.  
 XX  
 AC AAR74333;  
 XX  
 DT 01-NOV-1995 (first entry)  
 XX  
 DE Japanese cedar pollen allergen.  
 XX  
 DE Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;  
 KW desensitizer.  
 KW  
 XX Cryptomeria japonica.  
 OS  
 XX EP655500-A.  
 XX  
 XX 31-MAY-1995.  
 PD  
 XX 03-NOV-1994; 94EP-0308117.  
 PF  
 XX 27-DEC-1993; 93JP-0346814.  
 PR 05-NOV-1993; 93JP-0299151.  
 PR 20-DEC-1993; 93JP-0344596.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 XX Kurimoto M, Namba M, Torigoe K;  
 PI  
 XX WPI; 1995-195586/26.  
 DR N-PSDB; AAQ90156.  
 DR  
 XX New Japanese cedar pollen allergen polypeptide - and DNA coding for it,  
 PT useful for treatment and diagnosis of cedar pollen allergy  
 XX  
 XX Claim 5; Page 26-28; 41pp; English.  
 PS  
 CC The gene encoding an allergen of Japanese cedar pollen was isolated  
 CC by PCR amplification using primers based on portions of the allergen  
 CC protein. The gene was used for recombinant allergen production in  
 CC E. coli (vector plasmid pKK-223-3).  
 CC  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 82; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NNRIMLOFAKLTGFT 15  
 DB 135 NNRIMLOFAKLTGFT 149

```

RESULT 7
AAR69792 standard; Protein; 514 AA.
XX
XX
AC AAR69792;
XX
DT 27-SEP-1995 (first entry)
XX
DE Japonicum allergen.
XX
KW Japonicum allergen; induced histamine release; antiallergic peptide;
XX IGE cross-linking inhibition.
XX
OS Japonicum sp.
XX
PN W09502412-A.
XX
PD 26-JAN-1995.
XX
PF 15-JUL-1994; 94MO-JP01164.
XX
PR 16-JUL-1993; 93JP-0177008.
XX 01-SEP-1993; 93JP-0217725.
PR 07-APR-1994; 94JP-0069336.
XX
PA (MEIP ) MEIJI MILK PROD CO LTD.
XX
PI Kino K, Kohno Y, Komiyama N, Sone T;
XX
DR WPI; 1995-067159/09.
XX N-PSDB; AAO84045, AAO84046.
XX
PT Peptide anti-allergic agent - inhibits cross-linking of allergen
XX with IGE antibody
XX
PS Example 3; Pages 27-28; 46pp; Japanese.
XX
CC AAO84045 encodes AAR69792 Japonicum allergen, from which the
XX anti-allergic peptides AAR69845-R69809 were derived. The peptides ability
XX to inhibit the cross-linking of an allergen, to an IGE antibody can be
XX used in the prevention and treatment of allergic diseases.
XX
SQ Sequence 514 AA;
Query Match 100.0%; Score 82; DB 16; Length 514;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NNRIMLOFAKLTGFT 15
DB 135 NNRIMLOFAKLTGFT 149
RESULT 8
AAR3599 standard; Protein; 514 AA.
XX
XX
AC AAR3599;
XX
DT 16-AUG-1996 (first entry)
XX
DE Japan cedar pollen Cry j II allergen.
XX
KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;
XX Sugi pollinosis; diagnosis; treatment.
XX
OS Cryptomeria japonica.
XX
PI Key Location/Qualifiers
XX FH 1..54
XX FT /*label= sig_peptide
XX

```

```

FT Protein 55.514
XX /*label= mat_protein
XX
XX JP08047392-A.
XX
XX 20-FEB-1996.
XX
XX
PF 07-NOV-1994; 94JP-0297840.
XX
XX 26-MAY-1994; 94JP-0134868.
XX 05-NOV-1993; 93JP-0276773.
XX
XX
PA (MEIP ) MEIJI MILK PROD CO LTD.
XX
XX WPI; 1996-166249/17.
XX N-PSDB; AAT18102.
XX
XX Japan cedar pollen allergen Cry j II epitope - comprises at least
XX part of specified 460 amino acid protein
XX
XX Claim 1; Page 10-11; 17pp; Japanese.
XX
XX AAR3599 is a Japan cedar pollen Cry j II allergen which is useful
XX in the diagnosis, prevention and treatment of Sugi pollinosis,
XX the allergic reaction to Japan cedar pollen. Significant regions of
XX the allergen were identified using overlapping peptides of the full
XX epitope derived from a Cry j II antigen-specific T cell line
XX (see AAR97871-R97960). Amino acids 66-80 (AAR97884) and 186-200
XX (R978908) of the full mature 460 amino acid allergen are the most
XX allergenic of the 90 peptides tested.
XX
SQ Sequence 514 AA;
Query Match 100.0%; Score 82; DB 17; Length 514;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NNRIMLOFAKLTGFT 15
DB 135 NNRIMLOFAKLTGFT 149

```

```

RESULT 9
AAR1586 standard; Protein; 514 AA.
XX
XX
AC AAR1586;
XX
DT 24-MAY-1996 (first entry)
XX
DE Cedar pollen allergen A.
XX
XX Cedar; pollen; allergen; immunoglobulin E; IGE; T-cell epitope;
XX antibody; pollinosis; therapy; immunotherapy.
XX
XX Cryptomeria japonica.
XX
XX
XX EP700929-A2.
XX
XX 13-MAR-1996.
XX
XX 08-SEP-1995; 95EP-0306295.
XX
XX 14-JUL-1995; 95JP-0200221.
XX 10-SEP-1994; 94JP-0242137.
XX 14-JUL-1995; 95JP-0200204.
XX
XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.
XX
XX Hino K, Saito S, Taniguchi Y;
XX
XX WPI; 1996-140976/15.
XX

```

This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC class II molecule. The methods can be used for desensitising patients to allergens present in e.g., grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomids (non-biting midges), spiders and mites, housefly, fruit

This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of Tenbridio mottler beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence

CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II.

XX  
AC AAR81588;  
XX  
SQ Sequence 514 AA;

Query Match 100.0%; Score 82; DB 20; Length 514;  
Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NNRIMLOFAKLTGFT 15  
DB 135 NNRIMLOFAKLTGFT 149

RESULT 12  
AAR81581  
ID AAR81581 standard; Peptide; 17 AA.

XX  
AC AAR81581;  
XX  
DT 24-MAY-1996 (first entry)  
XX  
DE Cedar pollen allergen peptide 9 (T-cell epitope).  
XX  
DE Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;  
XX  
KW IGE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.  
XX  
OS Synthetic.  
XX  
PN EP700929-A2.  
XX  
PD 13-MAR-1996.  
XX  
PF 08-SEP-1995; 95EP-0306295.  
XX  
PR 14-JUL-1995; 95JP-0200221.  
XX  
PR 10-SEP-1994; 94JP-0242137.  
XX  
PR 14-JUL-1995; 95JP-0200204.  
XX  
PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
XX  
PI Hino K, Saito S, Taniguchi Y;  
XX  
DR WPI; 1996-140976/15.  
XX  
PT New peptide(s) derived from cedar pollen allergens - activate  
PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
PT used for treating cedar pollinosis  
XX  
PS Claim 4; Page 28; 36pp; English.  
XX  
CC Synthetic peptides based on portions of cedar pollen allergens A  
CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
CC cedar allergen-specific T-cells, but not allergen-specific IgE  
CC antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell  
CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
CC be used as immunotherapeutic agents to treat or prevent cedar  
CC pollinosis, avoiding side-effects such as anaphylaxis.  
XX  
SQ Sequence 17 AA;

Query Match 92.7%; Score 76; DB 17; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NNRIMLOFAKLTGFT 15  
DB 1 NNRIMLOFAKLTGFT 14

RESULT 13  
AAR81588  
ID AAR81588 standard; Peptide; 17 AA.

XX  
AC AAR81588;  
XX  
DT 24-MAY-1996 (first entry)  
XX  
DE Cedar pollen allergen peptide 16 (homologue).  
XX  
DE Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;  
XX  
KW IGE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.  
XX  
OS Synthetic.  
XX  
PN EP700929-A2.  
XX  
PD 13-MAR-1996.  
XX  
PF 08-SEP-1995; 95EP-0306295.  
XX  
PR 14-JUL-1995; 95JP-0200221.  
XX  
PR 10-SEP-1994; 94JP-0242137.  
XX  
PR 14-JUL-1995; 95JP-0200204.  
XX  
PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
XX  
PI Hino K, Saito S, Taniguchi Y;  
XX  
DR WPI; 1996-140976/15.  
XX  
PT New peptide(s) derived from cedar pollen allergens - activate  
PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
PT used for treating cedar pollinosis  
XX  
PS Disclosure; Page 32; 36pp; English.  
XX  
CC Synthetic peptides based on portions of cedar pollen allergens A  
CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
CC cedar allergen-specific T-cells, but not allergen-specific IgE  
CC antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell  
CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
CC for T-cell recognition, and homologous peptides (AAR81588-96)  
CC that contain 1 or more amino acid substitution(s) can be used  
CC as immunotherapeutic agents to treat or prevent cedar pollinosis,  
CC avoiding side-effects such as anaphylaxis.  
XX  
SQ Sequence 17 AA;

Query Match 85.4%; Score 70; DB 17; Length 17;  
Best Local Similarity 92.9%; Pred. No. 1.4e-05;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 NNRIMLOFAKLTGFT 15  
DB 1 NNRIMLOFAKLTGFT 14

RESULT 14  
AAR80346  
ID AAR80346 standard; peptide; 18 AA.

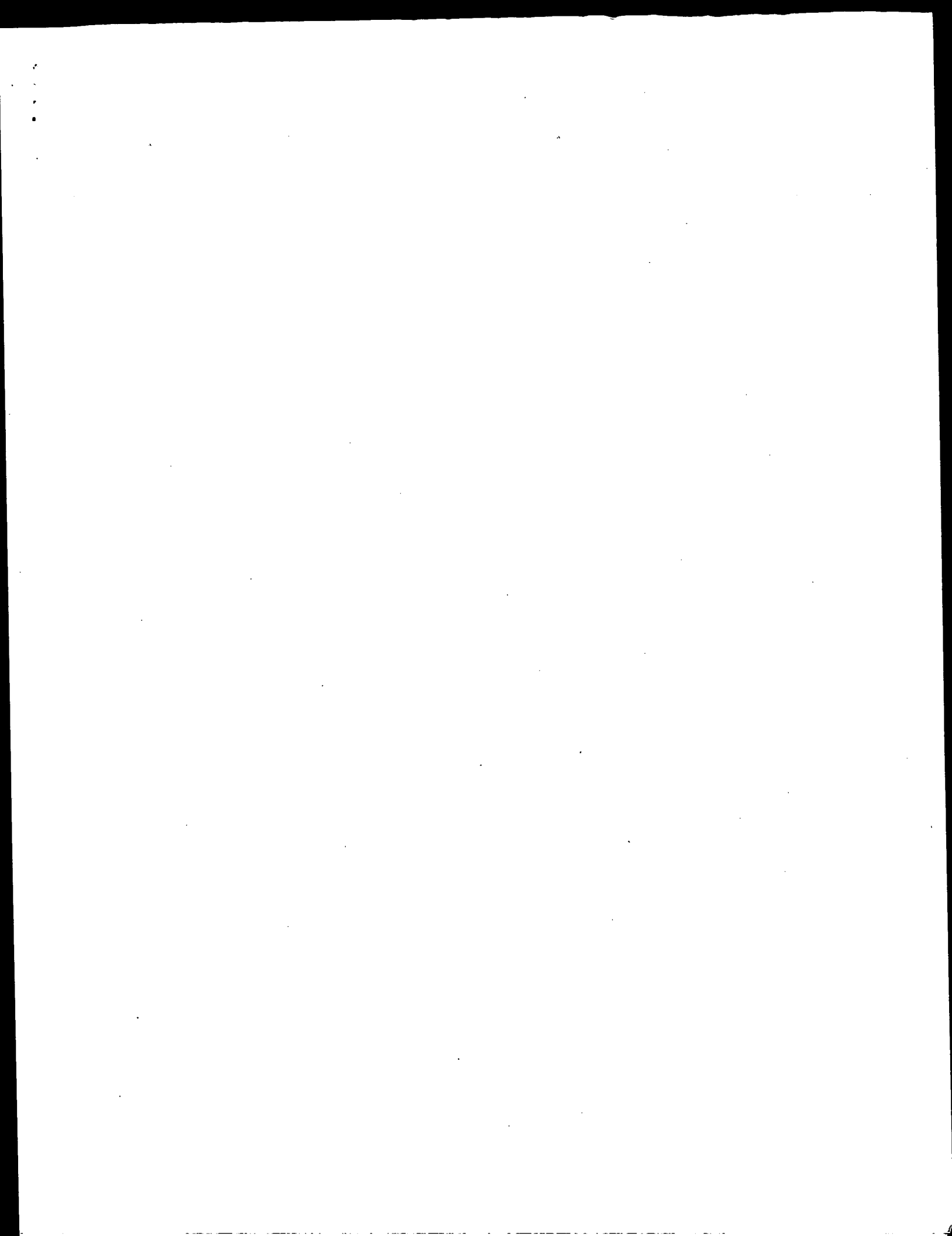
XX  
AC AAR80346;  
XX  
DT 11-JAN-1999 (first entry)  
XX  
DE Sugi allergen protein Cryj2 derived epitope for T cells.  
XX  
DE T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
XX  
KW sugi-pollinosis; allergic reaction; pollen.  
XX  
OS Synthetic.  
XX  
PN JP10259198-A.  
XX  
PD 29-SEP-1998.

XX 22-DEC-1997; 97JP-0353448.  
 XX 24-DEC-1996; 96JP-0343441.  
 XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY ) SANKYO CO LTD.  
 XX WPI; 1998-577037/49.  
 DR A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 XX Claim 7; Page 18; 21pp; Japanese.  
 CC AAW80339-58 represent epitopes for T cells, derived from the sugi  
 CC allergen proteins Cry1 (AAW80339-44, AAW80350-53 and AAW80356-58) and  
 CC Cry2 (AAW80345-49 and AAW80354-55). The peptides are useful for the  
 CC treatment of sugi-pollinosis, an allergic reaction of the body to  
 CC pollen.  
 SQ Sequence 18 AA;

Query Match 85.4%; Score 70; DB 19; Length 18;  
 Best Local Similarity 100.0%; Pred.No.1.5e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 RIMLOFAKLTGFT 15  
 DB 1 RIMLOFAKLTGFT 13

RESULT 15  
 AAW12542  
 ID AAW12542 standard; peptide; 13 AA.  
 AC AAW12542;  
 XX 30-APR-1997 (first entry)  
 DT Japanese cedar pollen allergen-derived peptide 11.  
 DE Japanese cedar pollen allergen-derived peptide 11.  
 XX Japanese cedar; sugi; pollen; pollinosis; pollenosis; allergen;  
 KW allergy; T-cell epitope; TCR; T-cell receptor; activation;  
 KM immune tolerance; Cryj.  
 XX Cryptomeria japonica.  
 OS Synthetic.  
 XX JP08333391-A.  
 PN 17-DEC-1996.  
 PD 18-JUL-1995; 95JP-0181438.  
 PF 07-APR-1995; 95JP-0082519.  
 PR (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY ) SANKYO CO LTD.  
 XX WPI; 1997-095487/09.  
 DR Peptide allergen derived from Japanese cedar pollen - causes T cell  
 PT response specific to cedar pollen, for treatment of pollenosis  
 XX Claim 16; Page 19; 21pp; Japanese.  
 PS The present sequence is one of 24 claimed peptides which were  
 CC synthesised based on Japanese cedar pollen sequences. This peptide  
 CC was shown to have Japanese cedar pollen antigen T cell epitope  
 CC activity by using T cells isolated from a cedar pollinosis patient.  
 CC The peptide produces little or no anaphylaxis.

SQ Sequence 13 AA;  
 Query Match 79.3%; Score 65; DB 18; Length 13;  
 Best Local Similarity 100.0%; Pred.No.8.2e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 IWLQFAKLTGFT 15  
 DB 1 IWLQFAKLTGFT 12  
 Search completed: April 20, 2003, 13:06:09  
 Job time : 18.1974 secs





Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds

(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524D-101

Sequence: 1 LQFAKL TGF TLMGKG 15

Scoring table: BLOSUM62

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SID82/gcgdata/genseq/genseqp-emb1/AA1980.DAT.*
2:	/SID82/gcgdata/genseq/genseqp-emb1/AA1981.DAT.*
3:	/SID82/gcgdata/genseq/genseqp-emb1/AA1982.DAT.*
4:	/SID82/gcgdata/genseq/genseqp-emb1/AA1983.DAT.*
5:	/SID82/gcgdata/genseq/genseqp-emb1/AA1984.DAT.*
6:	/SID82/gcgdata/genseq/genseqp-emb1/AA1985.DAT.*
7:	/SID82/gcgdata/genseq/genseqp-emb1/AA1986.DAT.*
8:	/SID82/gcgdata/genseq/genseqp-emb1/AA1987.DAT.*
9:	/SID82/gcgdata/genseq/genseqp-emb1/AA1988.DAT.*
10:	/SID82/gcgdata/genseq/genseqp-emb1/AA1989.DAT.*
11:	/SID82/gcgdata/genseq/genseqp-emb1/AA1990.DAT.*
12:	/SID82/gcgdata/genseq/genseqp-emb1/AA1991.DAT.*
13:	/SID82/gcgdata/genseq/genseqp-emb1/AA1992.DAT.*
14:	/SID82/gcgdata/genseq/genseqp-emb1/AA1993.DAT.*
15:	/SID82/gcgdata/genseq/genseqp-emb1/AA1994.DAT.*
16:	/SID82/gcgdata/genseq/genseqp-emb1/AA1995.DAT.*
17:	/SID82/gcgdata/genseq/genseqp-emb1/AA1996.DAT.*
18:	/SID82/gcgdata/genseq/genseqp-emb1/AA1997.DAT.*
19:	/SID82/gcgdata/genseq/genseqp-emb1/AA1998.DAT.*
20:	/SID82/gcgdata/genseq/genseqp-emb1/AA1999.DAT.*
21:	/SID82/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
22:	/SID82/gcgdata/genseq/genseqp-emb1/AA2001.DAT.*
23:	/SID82/gcgdata/genseq/genseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	76	100.0	15	17	AA979888	Japan cedar pollen
2	76	100.0	18	19	AAW80346	Sugi allergen prot
3	76	100.0	460	16	AA869791	Japonicum allergen
4	76	100.0	514	15	AA853690	Japonicum cedar pol
5	76	100.0	514	16	AA874333	Japanese cedar pol
6	76	100.0	514	16	AA869792	Japonicum allergen
7	76	100.0	514	17	AA835959	Japan cedar pollen
8	76	100.0	514	20	AA815866	Cedar pollen aller
9	76	100.0	514	17	AA825665	Japanese cedar all
10	76	100.0	514	20	AA825667	Japanese cedar all

1	67	88.2	105	18	AAW27370	Multi-epitope pep
2	67	88.2	134	18	AAW27371	Multi-epitope pep
3	65	85.5	14	18	AAW12535	Japanese cedar pol
4	65	85.5	17	17	AAAB1581	Cedar pollen aller
5	63	82.9	16	22	AAAB69104	Cedar pollen aller
6	63	82.9	96	22	AAAB69119	Cedar pollen aller
7	63	82.9	97	22	AAAB69105	Cedar pollen aller
8	63	82.9	210	22	AAAB69103	Cedar pollen aller
9	63	82.9	212	22	AAAB69120	Cedar pollen aller
10	61	80.3	12	18	AAW12557	Cedar pollen aller
11	61	80.3	12	22	AAAB69107	Japanese cedar pol
12	61	80.3	93	21	AAAB3891	Cryptomeria japoni
13	61	80.3	93	21	AAAB3896	Artificial sequenc
14	61	80.3	93	21	AAAB69097	Artificial sequenc
15	61	80.3	93	22	AAAB69117	Cedar pollen aller
16	61	80.3	95	21	AAAB3897	Cedar pollen aller
17	61	80.3	95	21	AAAB3901	Artificial sequenc
18	61	80.3	95	21	AAAB69088	Artificial sequenc
19	61	80.3	99	21	AAAB3902	Cedar pollen aller
20	61	80.3	99	21	AAAB23907	Artificial sequenc
21	61	80.3	99	22	AAAB69099	Artificial sequenc
22	61	80.3	12	18	AAW12554	Cedar pollen aller
23	59	77.6	13	18	AAW12556	Japanese cedar pol
24	59	77.6	17	17	AAAB1588	Japanese cedar pol
25	59	77.6	94	21	AAAB23906	Artificial sequenc
26	55	72.4	11	18	AAW12546	Japanese cedar pol
27	55	72.4	11	18	AAW12543	Japanese cedar pol
28	54	71.1	12	18	AAW12537	Japanese cedar pol
29	54	71.1	13	18	AAW12542	Japanese cedar pol
30	54	71.1	514	17	AAW04346	Chamaecyparis obtu
31	54	71.1	514	17	AAW042122	Japanese cypress p
32	54	71.1	15	17	AAAB97889	Japan cedar pollen
33	52	68.4	15	17	AAAB97889	Japanese cedar pol
34	52	68.4	15	18	AAW12545	Japanese cedar pol
35	50	65.8	15	17	AAAB97887	Japan cedar pollen
36	50	65.8	15	17	AAW121772	T-cell epitope pep
37	50	65.8	20	19	AAW421772	

## ALIGNMENTS

RESULT 1  
AAR97888

DT	16-AUG-1996	(first entry)
VV		

Japan cedar pollen mature allergen Cry j II amino acids 86-100.

KW Allergen; epitope; overlapping peptide; Cry 3 II; cedar pollen  
 KM Sugi pollinosis; diagnosis; treatment.  
 KY

*Cryptomeria japonica*.

PN JP08047392-A  
XY

20-FEB-1996  
PD  
XX

07-NOV-1994; 94JP-0297840.

PR 26-MAY-1994; 94JP-0134868.  
DE NOV 1993 03 11 00 00 00

XX  
XX  
(METE) METE

[illegible]

33

PT part of specified 460 amino acid protein

PS Claim 8; Fig 3; 17pp; Japanese.

AA697871-R97960 are overlapping peptides used for the epitope mapping of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic peptides of it are useful in the diagnosis, prevention and treatment of Sugi pollinosis, the allergic reaction to Japan cedar pollen. Significant regions of the allergen were identified using the overlapping peptides of the full epitope derived from a Cry j II antigen-specific T cell line. Amino acids 66-80 (AA697884) and 186-200 (R978908) of the full mature 460 amino acid allergen are the most allergenic of the 90 peptides tested.

Sequence 15 AA;

Query Match 100.0%; Score 76; DB 17; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LOPAKLTGFTLMGKG 15  
DB 1 LOPAKLTGFTLMGKG 15

RESULT 2

AA680346  
ID AA680346 standard; peptide; 18 AA.

AC AA680346;

DT 11-JAN-1999 (first entry)

DE Sugi allergen protein Cryj2 derived epitope for T cells.

KW T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment; sugi-pollinosis; allergic reaction; pollen.

OS Synthetic.

PN JP10259198-A.

PD 29-SEP-1998.

PF 22-DEC-1997; 97JP-0353448.

PR 24-DEC-1996; 96JP-0343441.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PA (SANY) SANKYO CO LTD.

DR WPI; 1998-577037/49.

PT A linked T cell epitope peptide - used for the treatment of sugi-pollinosis

PS Claim 7; Page 18; 21pp; Japanese.

AA680339-58 represent epitopes for T cells, derived from the sugi allergen proteins Cryj1 (AA680339-44, AA680350-53 and AA680356-58) and Cryj2 (AA680345-49 and AA680354-55). The peptides are useful for the treatment of sugi-pollinosis, an allergic reaction of the body to pollen.

Sequence 18 AA;

Query Match 100.0%; Score 76; DB 19; Length 18;  
Best Local Similarity 100.0%; Pred. No. 4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LOPAKLTGFTLMGKG 15  
DB 4 LOPAKLTGFTLMGKG 18

RESULT 3  
AA69791

ID AA69791 standard; Protein; 460 AA.

AC AA69791;

DT 27-SEP-1995 (first entry)

DE Japonicum allergen residues 55-514.

KW Japonicum allergen; residues 55-514; induced histamine release; anti-allergic peptide; IgE cross-linking inhibition.

OS Japonicum sp.

PN WO9502412-A.

PD 26-JAN-1995.

PF 15-JUL-1994; 94WO-JP01164.

PR 16-JUL-1993; 93JP-0177008.

PR 01-SEP-1993; 93JP-0217725.

PR 07-APR-1994; 94JP-0069336.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Kuno K, Kohno Y, Komiya N, Sone T;

DR WPI; 1995-067159/09.

DR N-PSDB; AA084044.

PT Peptide anti-allergic agent - inhibits cross-linking of allergen with IgE antibody

PS Disclosure; Pages 26-27; 46pp; Japanese.

AA084044 encodes AA69791 Japonicum allergen residues 55-514, from which the anti-allergic peptides AA69845-86980 were derived. The peptides ability to inhibit the cross-linking of an allergen, to an IgE antibody can be used in the prevention and treatment of allergic diseases.

Sequence 460 AA;

Query Match 100.0%; Score 76; DB 16; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LOPAKLTGFTLMGKG 15  
DB 86 LOPAKLTGFTLMGKG 100

RESULT 4

AA653690  
ID AA653690 standard; Protein; 514 AA.

AC AA653690;

DT 01-FEB-1995 (first entry)

DE Japanese cedar pollen allergen Cry j II.

KW Cedar pollinosis; diagnostic.

OS Cryptomeria japonica.

PN WO9411512-A.

PD 26-MAY-1994.

PF 12-NOV-1993; 93WO-US11000.

PR 12-NOV-1992; 92US-0975179.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX Brauer A, Kuo M, Pollock J, Yeung S;  
 XX WPI; 1994-183513/22.  
 DR N-PSDB; AAQ6048.  
 XX  
 PT Allergic Cry j II protein and fragments from Japanese cedar  
 PT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 XX  
 PS Claim 2; Fig 4; 89pp; English.  
 CC The sequence is of a Japanese cedar pollen allergen Cry j  
 CC II. The protein and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also AAR53692-6.  
 XX  
 SQ Sequence 514 AA;

Query Match 100.0%; Score 76; DB 15; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LOPAKLTGFTLMGKG 15  
 DB 140 LOPAKLTGFTLMGKG 154

RESULT 5  
 AAR74333  
 ID AAR74333 standard; Protein; 514 AA.  
 AC AAR74333;  
 XX  
 DT 01-NOV-1995 (first entry)  
 DE Japanese cedar pollen allergen.  
 DE Japanese cedar pollen allergen.  
 KM Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;  
 KM desensitizer.  
 XX Cryptomeria japonica.  
 OS  
 PN EP655500-A.  
 PD 31-MAY-1995.  
 XX  
 PF 03-NOV-1994; 94EP-0308117.  
 XX  
 PR 27-DEC-1993; 93JP-0346814.  
 PR 05-NOV-1993; 93JP-0299151.  
 PR 20-DEC-1993; 93JP-0344596.  
 XX  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX Kurimoto M, Namba M, Torigoe K;  
 PI WPI; 1995-195588/26.  
 DR N-PSDB; AAQ90156.  
 DR  
 XX  
 PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it,  
 PT useful for treatment and diagnosis of cedar pollen allergy  
 PS Claim 5; Page 26-28; 41pp; English.  
 CC The gene encoding an allergen of Japanese cedar pollen was isolated  
 CC by PCR amplification using primers based on portions of the allergen  
 CC protein. The gene was used for recombinant allergen production in  
 CC E. coli (vector plasmid pKK-223-3).  
 XX

SQ Sequence 514 AA;  
 Query Match 100.0%; Score 76; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LOPAKLTGFTLMGKG 15  
 DB 140 LOPAKLTGFTLMGKG 154

RESULT 6  
 AAR69792  
 ID AAR69792 standard; Protein; 514 AA.  
 AC AAR69792;  
 XX  
 DT 27-SEP-1995 (first entry)  
 DE Japonicum allergen.  
 DE Japonicum allergen.  
 KM Japonicum allergen; induced histamine release; antiallergic peptide;  
 KM IGE cross-linking inhibition.  
 XX Japonicum sp.  
 OS  
 PN W09502412-A.  
 PD 26-JAN-1995.  
 XX  
 PF 15-JUL-1994; 94WO-JP01164.  
 XX  
 PR 16-JUL-1993; 93JP-0177008.  
 PR 01-SEP-1993; 93JP-0217725.  
 PR 07-APR-1994; 94JP-0069336.  
 XX  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX  
 PI Kino K, Kohno Y, Komiyama N, Sone T;  
 XX  
 DR WPI; 1995-067159/09.  
 DR N-PSDB; AAQ84045; AAQ84046.  
 XX

Peptide antiallergic agent - inhibits cross-linking of allergen  
 with IGE antibody  
 PS Example 3; Pages 27-28; 46pp; Japanese.  
 CC AAQ84045 encodes AAR69792 Japonicum allergen, from which the  
 CC antiallergic peptides AAR69845-R69809 were derived. The peptides ability  
 CC to inhibit the cross-linking of an allergen, to an IGE antibody can be  
 CC used in the prevention and treatment of allergic diseases.  
 XX  
 SQ Sequence 514 AA;

Query Match 100.0%; Score 76; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LOPAKLTGFTLMGKG 15  
 DB 140 LOPAKLTGFTLMGKG 154

RESULT 7  
 AAR93599  
 ID AAR93599 standard; Protein; 514 AA.  
 AC AAR93599;  
 XX  
 DT 16-AUG-1996 (first entry)  
 DE Japan cedar pollen Cry j II allergen.  
 DE

XX Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
 KW Sugi pollinosis; diagnosis; treatment.  
 XX Cryptomeria japonica.  
 XX Key Location/Qualifiers  
 FT Peptide 1-54  
 FT /\*label= sig\_peptide  
 FT Protein 55..514  
 FT /\*label= mat\_protein  
 XX JP08047392-A.  
 PN 20-FEB-1996.  
 PD 07-NOV-1994; 94UP-0297840.  
 PF 26-MAY-1994; 94UP-0134868.  
 PR 05-NOV-1993; 93UP-0276773.  
 XX (MEIP ) MEIJI MILK PROD CO LTD.  
 PA WPI; 1996-166249/17.  
 DR N-PSDB; AAT18102.  
 XX Japan cedar pollen allergen Cry j II epitope - comprises at least  
 PT part of specified 460 amino acid protein  
 XX Claim 1; Page 10-11; 17pp; Japanese.  
 PS AAR93599 is a Japan cedar pollen Cry j II allergen which is useful  
 CC in the diagnosis, prevention and treatment of Sugi pollinosis,  
 CC the allergic reaction to Japan cedar pollen. Significant regions of  
 CC the allergen were identified using overlapping peptides of the full  
 CC epitope derived from a Cry j II antigen-specific T cell line  
 CC (see AAR97871-R97960). Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.  
 XX SQ Sequence 514 AA;  
 Query Match 100.0%; Score 76; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LQFAKLTGFTLMGKG 15  
 140 LQFAKLTGFTLMGKG 154  
 Db  
 RESULT 8  
 AAR81586 standard; Protein; 514 AA.  
 XX AAR81586;  
 AC 24-MAY-1996 (first entry)  
 DT Cedar pollen allergen A.  
 DE Cedar pollen allergen A.  
 XX Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KW antibody; pollinosis; therapy; immunotherapy.  
 XX Cryptomeria japonica.  
 OS EP700929-A2.  
 XX EP700929-A2.  
 PN 13-MAR-1996.  
 PD 08-SEP-1995; 95EP-0306295.  
 PF 14-JUL-1995; 95UP-0200221.  
 PR

PR 10-SEP-1994; 94UP-0242137.  
 PR 14-JUL-1995; 95UP-0200204.  
 XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX Hino K, Saito S, Taniguchi Y;  
 DR WPI; 1996-140976/15.  
 XX New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PT used for treating cedar pollinosis  
 XX Claim 5; Page 29-30; 36pp; English.  
 PS Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 XX SQ Sequence 514 AA;  
 Query Match 100.0%; Score 76; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LQFAKLTGFTLMGKG 15  
 140 LQFAKLTGFTLMGKG 154  
 Db  
 RESULT 9  
 AAY25666 standard; Protein; 514 AA.  
 XX AAY25666;  
 AC 30-SEP-1999 (first entry)  
 DT Japanese cedar allergen 1076242 Cry j II precursor protein fragment.  
 DE Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX Cedrus sp.  
 OS WO9934826-A1.  
 XX WO9934826-A1.  
 PN 15-JUL-1999.  
 PD 11-JAN-1999; 99WO-GB00080.  
 PF 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PA Kay AB, Larche M;  
 XX Kay AB, Larche M;  
 PI WPI; 1999-458255/38.  
 DR Desensitizing patients to polypeptide allergens  
 PT Example 6; Page 74; 117pp; English.  
 PS

CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076242 Cry j II  
 CC precursor.

CC Sequence 514 AA;

Query Match 100.0%; Score 76; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQPAKLTGFTLMGKG 15  
 DB 140 LQPAKLTGFTLMGKG 154

RESULT 10

AAV25667  
 ID AAV25667 standard; protein; 514 AA.

AC AAV25667;

DT 30-SEP-1999 (first entry)

XX Japanese cedar allergen 1076241 Cry j II protein fragment.

KW Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; mould; food; insect; stinging;  
 KW chironomidae; spider; mite; housefly; fruit fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

OS Cedrus sp.

XX WO9934826-A1.

PN 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

DR WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

PS Example 6; Page 74; 117pp; English.

CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for

CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II.

CC Sequence 514 AA;

Query Match 100.0%; Score 76; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQPAKLTGFTLMGKG 15  
 DB 140 LQPAKLTGFTLMGKG 154

RESULT 11

AAW27370  
 ID AAW27370 standard; peptide; 105 AA.

AC AAW27370;

DT 24-MAR-1998 (first entry)

XX Multi-epitope peptide used as immunotherapeutic agent #2.

KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

XX WO9732600-A1.

XX 12-SEP-1997.

XX 10-MAR-1997; 97WO-JP00740.

XX 10-MAR-1996; 96JP-0080702.

XX (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Iwama A, Kuno K, Kume A, Sone T;

DR WPI; 1997-470495/43.

XX Peptide immuno-therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens

XX Claim 6; Page 31; 58pp; Japanese.

CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

XX Sequence 105 AA;

Query Match 88.2%; Score 67; DB 18; Length 105;  
 Best Local Similarity 92.9%; Pred. No. 0.00013;

Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGK 14  
 |||||  
 DB 59 LQFAKLTGFTLMGR 72

RESULT 12  
 AAW27371  
 ID AAW27371 standard; peptide; 134 AA.

XX AAW27371;

DT 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #3.

KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

XX Synthetic.

FN WO9732600-A1.

PD 12-SEP-1997.

PF 10-MAR-1997; 97WO-JP00740.

PR 10-MAR-1996; 96JP-0080702.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;

DR WPI; 1997-470495/43.

PT Peptide immunotherapeutic agent to treat allergic diseases -  
 contains multi-epitope peptide containing T cell epitope regions

PT from different allergens

PS Claim 6; Page 32; 58pp; Japanese.

XX The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

XX Sequence 134 AA;

Query Match 88.2%; Score 67; DB 18; Length 134;

Best Local Similarity 92.9%; Pred. No. 0.00017;

Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMGK 14  
 |||||  
 DB 59 LQFAKLTGFTLMGR 72

RESULT 13  
 AAW12535  
 ID AAW12535 standard; peptide; 14 AA.

XX AAW12535;

DT 30-APR-1997 (first entry)

DE Japanese cedar pollen allergen-derived peptide 4.

KW Japanese cedar; sugi; pollen; pollinosis; pollenosis; allergen;  
 KW allergy; T-cell epitope; TCR; T-cell receptor; activation;  
 KW immune tolerance; CryJ).

OS Crytomeria japonica.

PN DP08333391-A.

PD 17-DEC-1996.

PE 18-JUL-1995; 95JP-0181438.

PR 07-APR-1995; 95JP-0082519.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

DR WPI; 1997-095487/09.

PT Peptide allergen derived from Japanese cedar pollen - causes T cell  
 PT response specific to cedar pollen, for treatment of pollinosis

PS Claim 5; Page 19; 21pp; Japanese.

XX The present sequence is one of 24 claimed peptides which were  
 CC synthesised based on Japanese cedar pollen sequences. This peptide  
 CC was shown to have Japanese cedar pollen antigen T cell epitope  
 CC activity by using T cells isolated from a cedar pollinosis patient.  
 CC The peptide produces little or no anaphylaxis.

XX Sequence 14 AA;

Query Match 85.5%; Score 65; DB 18; Length 14;

Best Local Similarity 100.0%; Pred. No. 3.2e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMG 13  
 |||||  
 DB 2 LQFAKLTGFTLMG 14

RESULT 14  
 AAR81581  
 ID AAR81581 standard; peptide; 17 AA.

XX AAR81581;

DT 24-MAY-1996 (first entry)

DE Cedar pollen allergen peptide 9 (T-cell epitope).

KW Cedar; Crytomeria japonica; pollen; allergen; immunoglobulin E;  
 KW IgE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.

XX Synthetic.

PN EP700929-A2.

PD 13-MAR-1996.

PE 08-SEP-1995; 95EP-0306295.

PR 14-JUL-1995; 95JP-0200221.

PR 10-SEP-1994; 94JP-0242137.

PR 14-JUL-1995; 95JP-0200204.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Hino K, Saito S, Taniguchi Y;

DR WPI; 1996-140976/15.

PT New peptide(s) derived from cedar pollen allergens - activate  
PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
used for treating cedar pollinosis

XX Claim 4; Page 28; 36pp; English.

CC Synthetic peptides based on portions of cedar pollen allergens A  
CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
CC cedar pollen-specific T-cells, but not allergen-specific IgE  
CC antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell  
CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
CC be used as immunotherapeutic agents to treat or prevent cedar  
CC pollinosis, avoiding side-effects such as anaphylaxis.

XX Sequence 17 AA;

Query Match 85.5%; Score 65; DB 17; Length 17;  
Best Local Similarity 100.0%; Pred. No. 3.9e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQFAKLTGFTLMG 13  
DB 5 LQFAKLTGFTLMG 17

RESULT 15

AAB69104  
ID AAB69104 standard; Protein; 96 AA.

XX AAB69104;

DT 23-APR-2001 (first entry)

DE Cedar pollen allergen T cell epitope derived protein SEQ ID NO:13.

XX Japanese cedar; Cryptomeria japonica; cedar pollen allergen;  
XX T cell epitope; anti-sugipollinosis.

OS Cryptomeria japonica.  
OS Synthetic.

XX JP2000327699-A.

PD 28-NOV-2000.

PF 15-MAR-2000; 2000JP-0071710.

XX 15-MAR-1999; 99JP-0068316.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
PA (SANY) SANKYO CO LTD.

XX WPI; 2001-185061/19.

XX Novel peptide and its use -

PS Claim 8; Page 40; 75pp; Japanese.

CC The present invention describes a peptide, its complex, derivative or  
CC its polymerizate, where the peptide (I) has a formula of:  
CC alpha1-alpha2-alpha3-alpha4-alpha5-alpha6-alpha7  
CC where alpha1 to alpha7 = amino acid sequence selected from amino acid  
CC sequences ranging from 11-19 amino acids derived from T cell epitopes  
CC derived from cedar (Japanese cedar - Cryptomeria japonica) pollen  
CC allergens. The peptide can be used in an anti-sugipollinosis agent.  
CC AAF59011 to AAF59062 and AAB69094 to AAB69121 represent sequences used  
CC in the exemplification of the present invention.

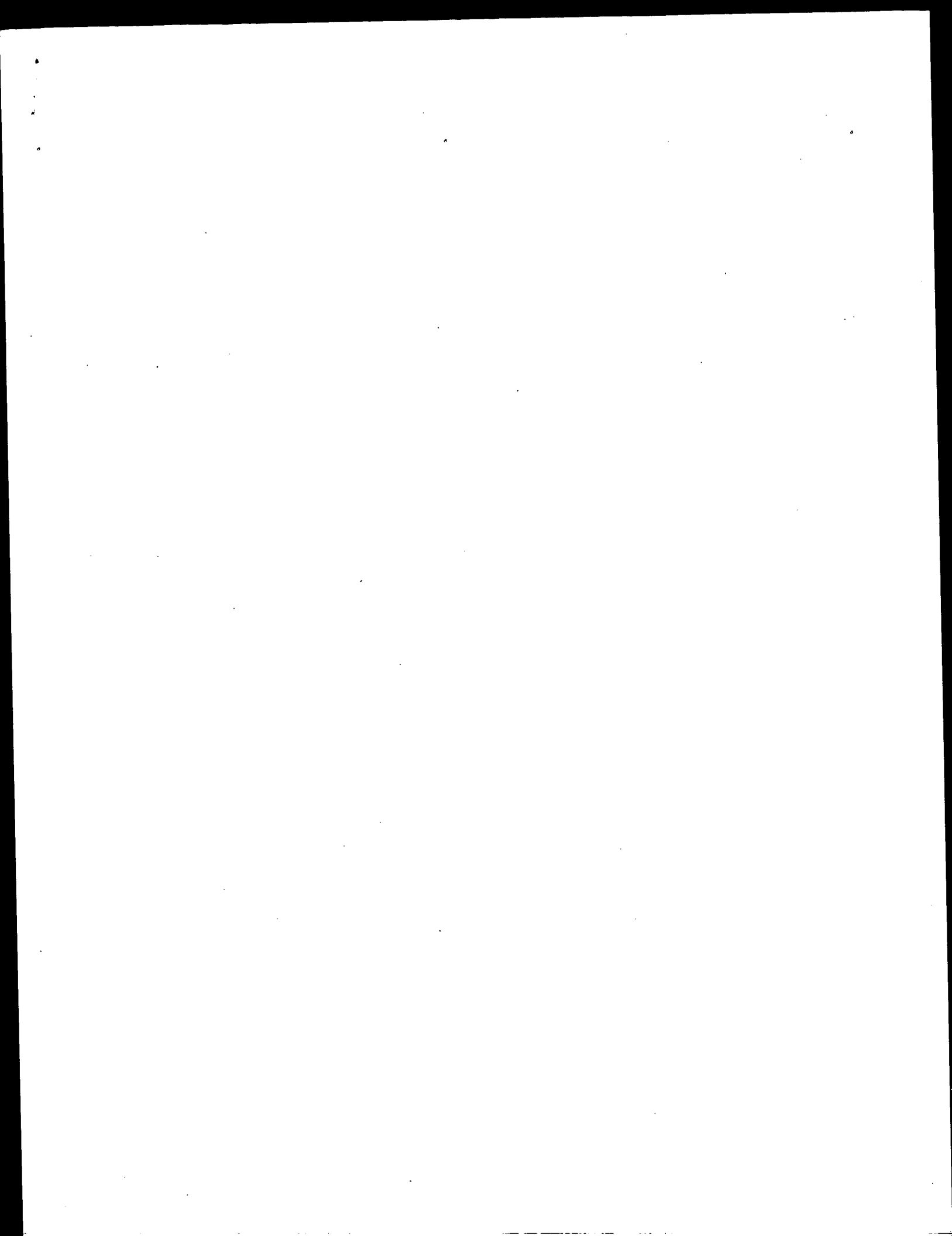
XX Sequence 96 AA;

Query Match 82.9%; Score 63; DB 22; Length 96;  
Best Local Similarity 92.3%; Pred. No. 0.00065;

Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QFAKLTGFTLMGX 14  
DB 84 QFAKLTGFTLMGR 96

Search completed: April 20, 2003, 13:06:10  
Job time: 19.1974 secs





GenCore version 5.1.4\_P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524d-120

Sequence: 1 GIDIFASKNFHLQKN 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	79 100.0	15 17	AA897907	Japan cedar pollen
2	79 100.0	15 19	AA857763	Residues 181-195 o
3	79 100.0	460 16	AA869791	Japonicum allergen
4	79 100.0	514 16	AA853690	Japanese cedar pol
5	79 100.0	514 16	AA874333	Japanese cedar pol
6	79 100.0	514 16	AA869792	Japonicum allergen
7	79 100.0	514 17	AA835599	Japan cedar pollen
8	79 100.0	514 17	AA81586	Cedar pollen aller
9	79 100.0	514 20	AA85566	Japanese cedar all
10	79 100.0	514 20	AA85567	Japanese cedar all

11	73 92.4	17 19	AAW80347	Sugi allergen prot
12	73 92.4	19 22	AAW84115	Amino acid sequenc
13	73 92.4	42 22	AAW84105	Immunomodulatory p
14	73 92.4	80 18	AAW27369	Multi-epitope pep
15	73 92.4	105 18	AAW27370	Multi-epitope pep
16	73 92.4	134 18	AAW27371	Multi-epitope pep
17	72 91.1	74 21	AAW23905	Artificial sequenc
18	72 91.1	94 21	AAW23906	Artificial sequenc
19	72 91.1	95 21	AAW23897	Artificial sequenc
20	72 91.1	95 21	AAW23901	Artificial sequenc
21	72 91.1	95 22	AAW60998	Artificial sequenc
22	72 91.1	96 22	AAW6104	Cedar pollen aller
23	72 91.1	96 22	AAW69109	Cedar pollen aller
24	72 91.1	97 22	AAW69105	Cedar pollen aller
25	72 91.1	99 21	AAW23902	Cedar pollen aller
26	72 91.1	99 21	AAW23907	Artificial sequenc
27	72 91.1	99 22	AAW60999	Artificial sequenc
28	72 91.1	214 22	AAW69120	Cedar pollen aller
29	69 87.3	13 22	AAW69101	Cryptomeria japoni
30	69 87.3	17 17	AAW81582	Japanese cedar pol
31	63 79.7	12 18	AAW12548	T-cell epitope pep
32	62 78.5	20 19	AAW42181	Chamaecyparis obtu
33	62 78.5	514 17	AAW04346	Japanese cypress p
34	62 78.5	514 19	AAW42122	Sugi allergen prot
35	61 77.2	33 19	AAW80340	Sugi allergen prot
36	61 77.2	33 19	AAW80341	Sugi allergen prot
37	59 74.7	11 18	AAW12550	Japanese cedar pol
38	59 74.7	12 18	AAW12549	Japanese cedar pol
39	59 74.7	12 22	AAW69108	Cryptomeria japoni
40	59 74.7	33 19	AAW60339	Sugi allergen prot
41	59 74.7	35 19	AAW80342	Sugi allergen prot
42	59 74.7	35 19	AAW80343	Sugi allergen prot
43	59 74.7	35 19	AAW80344	Sugi allergen prot
44	59 74.7	47 19	AAW80350	Sugi allergen prot
45	59 74.7	47 19	AAW80351	Sugi allergen prot

## ALIGNMENTS

RESULT 1  
AA897907  
ID AA897907 standard; peptide; 15 AA.  
XX  
AC AA897907;  
XX  
DT 16-AUG-1996 (first entry)  
XX  
DE Japan cedar pollen mature allergen Cry j II amino acids 181-195.  
XX  
KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
KW Sugi pollinosis; diagnosis; treatment.  
XX  
OS Cryptomeria japonica.  
XX  
PN JP08047392-A.  
XX  
PD 20-FEB-1996.  
XX  
PF 07-NOV-1994; 94UP-0297840.  
XX  
PR 26-MAY-1994; 94UP-0134868.  
XX  
PR 05-NOV-1993; 93UP-0276773.  
XX  
PA (MEIP) MEIJI MILK PROD CO LTD.  
XX  
WI 1996-166249/17.  
XX  
PT Japan cedar pollen allergen Cry j II epitope - comprises at least  
XX part of specified 460 amino acid protein  
XX  
PS Claim 8; Fig 4; 17pp; Japanese.  
XX

CC AAR97871-R97960 are overlapping peptides used for the epitope mapping  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.

CC Sequence 15 AA;

Query Match 100.0%; Score 79; DB 17; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLOKN 15  
 DB 1 GIDIFASKNFHLOKN 15

RESULT 2  
 AAM57763  
 ID AAM57763 standard; peptide; 15 AA.

XX AAM57763;

DT 17-SEP-1998 (first entry)

DE Residues 181-195 of Cry j 2.

XX Cry j 2; Japanese cedar pollen antigen; allergy; immunotherapy;

KW HLA class II molecule.

OS Cryptomeria japonica.

XX WO9820902-A1.

PD 22-MAY-1998.

PF 12-NOV-1997; 97WO-JP04129.

PR 13-NOV-1996; 96JP-0302053.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kino K, Kume A, Sone T;

XX WPI; 1998-297617/26.

PT Peptides derived from Japanese cedar pollen antigens are  
 PT immunotherapeutic agents - useful for allergy treatment and typing

PT HLA class II molecules in allergy sufferers

PS Claim 12; Page 31; 50pp; Japanese.

XX This sequence represents residues 181-195 of the Cry j 2 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.

XX Sequence 15 AA;

Query Match 100.0%; Score 79; DB 19; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIDIFASKNFHLOKN 15  
 DB 1 GIDIFASKNFHLOKN 15

RESULT 3  
 AAR69791  
 ID AAR69791 standard; Protein; 460 AA.

XX AAR69791;

DT 27-SEP-1995 (first entry)

DE Japonicum allergen residues 55-514.

XX Japonicum allergen; residues 55-514; induced histamine release;

KW antiallergic peptide; IgE cross-linking inhibition.

OS Japonicum sp.

XX WO9502412-A.

PD 26-JAN-1995.

PF 15-JUL-1994; 94WO-JP01164.

PR 16-JUL-1993; 93JP-0177008.

PR 01-SEP-1993; 93JP-0217725.

PR 07-APR-1994; 94JP-0069336.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Kino K, Kohno Y, Komiyama N, Sone T;

XX WPI; 1995-067159/09.

DR N-PSDB; AAO84044.

XX Peptide antiallergic agent - inhibits cross-linking of allergen

PT with IgE antibody

PS Disclosure; Pages 26-27; 46pp; Japanese.

XX AAO84044 encodes AAR69791 Japonicum allergen residues 55-514, from

CC which the antiallergic peptides AAR69845-R69809 were derived.

CC The peptides ability to inhibit the cross-linking of an allergen.

CC to an IgE antibody can be used in the prevention and treatment of

CC allergic diseases.

XX Sequence 460 AA;

QY 1 GIDIFASKNFHLOKN 15  
 DB 181 GIDIFASKNFHLOKN 195

RESULT 4  
 AAR53690  
 ID AAR53690 standard; Protein; 514 AA.

XX AAR53690;

DT 01-FEB-1995 (first entry)

DE Japanese cedar pollen allergen Cry j II.

XX Cedar pollinosis; diagnostic.

OS Cryptomeria japonica.

PN W09411512-A.  
 XX 26-MAY-1994.  
 PD  
 XX 12-NOV-1993; 93WO-US11000.  
 PF  
 XX 12-NOV-1992; 92US-0975179.  
 PR  
 XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PA  
 XX Brauer A, Kuo M, Pollock J, Yeung S;  
 PI  
 XX WPI; 1994-183513/22.  
 DR  
 XX N-PSDB; AAQ65048.  
 DR  
 XX Allergic Cry j II protein and fragments from Japanese cedar  
 PT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 XX  
 PS Claim 2; Fig 4; 89pp; English.  
 XX  
 CC The sequence is of a Japanese cedar pollen allergen Cry j  
 CC II. The protein and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also AAR33692-6.  
 CC  
 XX  
 SQ Sequence 514 AA;  
 XX  
 Query Match 100.0%; Score 79; DB 15; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GIDIFASKNFHLQKN 15  
 |||||  
 Db 235 GIDIFASKNFHLQKN 249  
 |||||  
 RESULT 5  
 AAR74333  
 ID AAR74333 standard; Protein; 514 AA.  
 XX  
 AC AAR74333;  
 XX  
 XX 01-NOV-1995 (first entry)  
 DT  
 XX Japanese cedar pollen allergen.  
 DE  
 XX Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;  
 KW desensitizer.  
 KM  
 XX Cryptomeria japonica.  
 OS  
 XX EP655500-A.  
 PN  
 XX 31-MAY-1995.  
 PD  
 XX 03-NOV-1994; 94EP-0308117.  
 PF  
 XX 27-DEC-1993; 93JP-0346814.  
 PR  
 XX 05-NOV-1993; 93JP-0299151.  
 PR  
 XX 20-DEC-1993; 93JP-0344596.  
 PR  
 XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA  
 XX Kurimoto M, Namba M, Torigoe K;  
 PI  
 XX WPI; 1995-195588/26.  
 DR  
 XX N-PSDB; AAQ90156.  
 DR  
 XX New Japanese cedar pollen allergen polypeptide - and DNA coding for it,  
 PT useful for treatment and diagnosis of cedar pollen allergy  
 PT  
 XX

PS Claim 5; Page 26-28; 41pp; English.  
 XX  
 CC The gene encoding an allergen of Japanese cedar pollen was isolated  
 CC by PCR amplification using primers based on portions of the allergen  
 CC protein. The gene was used for recombinant allergen production in  
 CC E. coli (vector plasmid pKK-223-3).  
 CC  
 XX  
 SQ Sequence 514 AA;  
 XX  
 Query Match 100.0%; Score 79; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GIDIFASKNFHLQKN 15  
 |||||  
 Db 235 GIDIFASKNFHLQKN 249  
 |||||  
 RESULT 6  
 AAR69792  
 ID AAR69792 standard; Protein; 514 AA.  
 XX  
 AC AAR69792;  
 XX  
 DT 27-SEP-1995 (first entry)  
 DT  
 XX Japonicum allergen.  
 DE  
 XX Japonicum allergen; induced histamine release; antiallergic peptide;  
 KW IGE cross-linking inhibition.  
 KW  
 XX Japonicum sp.  
 OS  
 XX W09502412-A.  
 PN  
 XX 26-JAN-1995.  
 PD  
 XX 15-JUL-1994; 94WO-JP01164.  
 PF  
 XX 16-JUL-1993; 93JP-0177008.  
 PR  
 XX 01-SEP-1993; 93JP-0217725.  
 PR  
 XX 07-APR-1994; 94JP-0069336.  
 PR  
 XX (MEIP) MEIJI MILK PROD CO LTD.  
 PA  
 XX  
 XX Kino K, Kohno Y, Komiyama N, Sone T;  
 PI  
 XX WPI; 1995-067159/09.  
 DR  
 XX N-PSDB; AAQ84045; AAQ84046.  
 DR  
 XX Peptide antiallergic agent - inhibits cross-linking of allergen  
 PT with IGE antibody  
 PT  
 XX Example 3; Pages 27-28; 46pp; Japanese.  
 PS  
 XX AAQ84045 encodes AAR69792 Japonicum allergen, from which the  
 CC antiallergic peptides AAR69845-R69809 were derived. The peptides ability  
 CC to inhibit the cross-linking of an allergen, to an IGE antibody can be  
 CC used in the prevention and treatment of allergic diseases.  
 CC  
 XX  
 SQ Sequence 514 AA;  
 XX  
 Query Match 100.0%; Score 79; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GIDIFASKNFHLQKN 15  
 |||||  
 Db 235 GIDIFASKNFHLQKN 249  
 |||||  
 RESULT 7  
 AAR33599

ID AAR93599 standard; Protein; 514 AA.  
 AC AAR93599;  
 XX  
 DT 16-AUG-1996 (first entry)  
 DE Japan cedar pollen Cry j II allergen.  
 XX  
 KM Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
 KM Sugi pollinosis; diagnosis; treatment.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..54  
 FT /\*label= sig\_peptide  
 FT Protein 55..514  
 FT /\*label= mat\_protein  
 XX  
 PN JP08047392-A.  
 PD 20-FEB-1996.  
 PF 07-NOV-1994; 94JP-0297840.  
 PR 26-MAY-1994; 94JP-0134868.  
 PR 05-NOV-1993; 93JP-0276773.  
 XX  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 XX  
 DR MPI; 1996-166249/17.  
 DR N-PSDB; AAT18102.  
 XX  
 PT Japan cedar pollen allergen Cry j II epitope - comprises at least  
 PT part of specified 460 amino acid protein  
 XX  
 PS Claim 1; Page 10-11; 17pp; Japanese.  
 XX  
 CC AAR93599 is a Japan cedar pollen Cry j II allergen which is useful  
 CC in the diagnosis, prevention and treatment of Sugi pollinosis;  
 CC the allergic reaction to Japan cedar pollen. Significant regions of  
 CC the allergen were identified using overlapping peptides of the full  
 CC epitope derived from a Cry j II antigen-specific T cell line  
 CC (see AAR97871-R97960). Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.  
 CC  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 79; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GIDIFASKNFHLQKN 15  
 DB 235 GIDIFASKNFHLQKN 249  
 RESULT 8  
 AAR81586  
 ID AAR81586 standard; Protein; 514 AA.  
 AC AAR81586;  
 XX  
 DT 24-MAY-1996 (first entry)  
 DE Cedar pollen allergen A.  
 XX  
 KM Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KM antibody; pollinosis; therapy; immunotherapy.  
 XX  
 OS Cryptomeria japonica.  
 XX

PN EP700929-A2.  
 PD 13-MAR-1996.  
 XX  
 PF 08-SEP-1995; 95SEP-0306295.  
 PR 14-JUL-1995; 95JP-0200221.  
 PR 10-SEP-1994; 94JP-0242137.  
 PR 14-JUL-1995; 95JP-0200204.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI Hino K, Saito S, Taniguchi Y;  
 XX  
 DR MPI; 1996-140976/15.  
 PT New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PT used for treating cedar pollinosis  
 XX  
 PS Claim 5; Page 29-30; 36pp; English.  
 XX  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 CC  
 SQ Sequence 514 AA;  
 Query Match 100.0%; Score 79; DB 17; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GIDIFASKNFHLQKN 15  
 DB 235 GIDIFASKNFHLQKN 249  
 RESULT 9  
 AAY25666  
 ID AAY25666 standard; Protein; 514 AA.  
 AC AAY25666;  
 XX  
 DT 30-SEP-1999 (first entry)  
 DE Japanese cedar allergen 1076242 Cry j II precursor protein fragment.  
 XX  
 KM Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mold; food; insect; sting;  
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larva; mealworm; cat;  
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 XX  
 PN WO9934826-A1.  
 PD 15-JUL-1999.  
 PF 11-JAN-1999; 99WO-GB00080.  
 PR 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX  
 PI Kay AB, Larche M;

XX WPI, 1999-458255/38.  
 XX Desensitizing patients to polypeptide allergens  
 XX Example 6; Page 74; 117pp; English.  
 PS  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC Tenbrio molitor beetle, bee moth larvae, mealworm, cockroach, larvae of  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076242 Cry j II  
 CC precursor.  
 CC  
 CC Sequence 514 AA;  
 SQ  
 Query Match 100.0%; Score 79; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 GIDIFASKNFHLOKN 15  
 Db 235 GIDIFASKNFHLOKN 249  
 RESULT 10  
 AAY25667  
 ID AAY25667 standard; protein; 514 AA.  
 AC  
 AC AAY25667;  
 DT 30-SEP-1999 (first entry)  
 DE Japanese cedar allergen 1076241 Cry j II protein fragment.  
 DE  
 XX Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;  
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KM cockroach; beetle; dog; horse; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 OS  
 PN WO9934826-A1.  
 PN  
 XX 15-JUL-1999.  
 PD  
 PD 11-JAN-1999; 99WC-GB00080.  
 PF  
 PF 21-SEP-1998; 98GB-0020474.  
 PR  
 PR 09-JAN-1998; 98GB-0000445.  
 XX  
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PA  
 PI Kay AB, Larche M;  
 PI  
 XX WPI, 1999-458255/38.  
 XX  
 XX Desensitizing patients to polypeptide allergens  
 XX Example 6; Page 74; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC Tenbrio molitor beetle, bee moth larvae, mealworm, cockroach, larvae of  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II.  
 CC  
 CC Sequence 514 AA;  
 SQ  
 Query Match 100.0%; Score 79; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 GIDIFASKNFHLOKN 15  
 Db 235 GIDIFASKNFHLOKN 249  
 RESULT 11  
 AAM80347  
 ID AAM80347 standard; peptide; 17 AA.  
 AC  
 AC AAM80347;  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj2 derived epitope for T cells.  
 DE  
 XX T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
 KM sugi-pollinosis; allergic reaction; pollen.  
 KM  
 XX Synthetic.  
 OS  
 OS JP10259198-A.  
 PN  
 PN 29-SEP-1998.  
 PD  
 PD 22-DEC-1997; 97JP-0353448.  
 PF  
 PF 24-DEC-1996; 96JP-0343441.  
 PR  
 PR (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA  
 PA (SANT) SANKYO CO LTD.  
 XX  
 XX WPI, 1998-577037/49.  
 XX  
 XX A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 PT  
 PT Claim 7; Page 18; 21pp; Japanese.  
 PS  
 PS AAM80339-58 represent epitopes for T cells, derived from the sugi  
 CC allergen proteins Cryj1 (AAM80339-44, AAM80350-53 and AAM80356-58) and  
 CC Cryj2 (AAM80345-49 and AAM80354-55). The peptides are useful for the  
 CC treatment of sugi-pollinosis, an allergic reaction of the body to  
 CC pollen.  
 CC  
 CC Sequence 17 AA;  
 SQ  
 Query Match 92.4%; Score 73; DB 19; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 IDIFASKNFHLQKN 15  
 DB 1 IDIFASKNFHLQKN 14

## RESULT 12

AAB84115  
 ID AAB84115 standard; peptide; 19 AA.

XX AAB84115;

DT 06-AUG-2001 (first entry)

DE Amino acid sequence of a pollen antigen.

KW Immunomodulatory peptide; immune response; autoimmune disease; allergy;  
 asthma; host-versus-graft rejection; T cell; anergy; apoptosis.

OS Cryptomeria japonica.

PN WO200136448-A2.

PD 25-MAY-2001.

PF 27-OCT-2000; 2000WO-US41646.

PR 27-OCT-1999; 99US-0161734.

PA (CELS-) CEL-SCI CORP.

PI Zimmerman DH;

DR WPI; 2001-374498/39.

PT Novel immunomodulatory peptide construct useful for modulating an  
 inappropriate immune response in an individual at risk for autoimmune  
 disease, allergic reactions, asthma or host-graft or graft-host disease

PS Claim 2; Page 27; 55pp; English.

XX The specification describes an immunomodulatory peptide construct.  
 CC The immunomodulatory peptide comprises a first peptide associated with  
 CC autoimmune disease, allergy or asthma, or host-versus-graft rejection  
 CC and which will bind to an antigen receptor on a set or subset of T cells,  
 CC linked to a second immune modulating peptide which will cause a directed  
 CC immune response by the set or subset of T cells to which the first  
 CC peptide is attached. Alternatively, the second peptide will bind to a  
 CC T cell receptor site on the surface of the T cell which will cause the  
 CC set or subset of T cells to initiate, but not complete, an immune  
 CC response to cause the set or subset of T cells to undergo anergy and  
 CC apoptosis. The immunomodulatory peptides are useful for eliminating a set  
 CC or subset of T cells involved in autoimmune response. They are useful for  
 CC the treatment of autoimmune disease, allergic reactions, asthma or  
 CC host-graft or graft-host rejection. The immunomodulatory peptides are  
 CC also useful for interrupting an autoimmune disease associated pathway  
 CC necessary to complete T cell activation. The present peptide is used to  
 CC construct immunomodulatory peptides of the invention, and is  
 CC representative of the first peptide.

XX Sequence 19 AA;

SO Query Match 92.4%; Score 73; DB 22; Length 19;

Best Local Similarity 100.0%; Pred. No. 1e-06; Mismatches 0; Indels 0; Gaps 0;

OY 2 IDIFASKNFHLQKN 15  
 DB 1 IDIFASKNFHLQKN 14

RESULT 13

AAB84105  
 ID AAB84105 standard; peptide; 42 AA.

XX AAB84105;

DT 06-AUG-2001 (first entry)

DE Immunomodulatory peptide used to treat allergic responses.

KW Immunomodulatory peptide; immune response; autoimmune disease; allergy;  
 asthma; host-versus-graft rejection; T cell; anergy; apoptosis.

OS Synthetic.

OS Cryptomeria japonica.

FT Key Location/Qualifiers

FT Peptide 1..17 /note= "LFA-3 peptide"

FT Peptide 18..23 /note= "linker"

FT Peptide 24..42 /note= "pollen antigen"

PN WO200136448-A2.

PD 25-MAY-2001.

PF 27-OCT-2000; 2000WO-US41646.

PR 27-OCT-1999; 99US-0161734.

PA (CELS-) CEL-SCI CORP.

PI Zimmerman DH;

DR WPI; 2001-374498/39.

PT Novel immunomodulatory peptide construct useful for modulating an  
 inappropriate immune response in an individual at risk for autoimmune  
 disease, allergic reactions, asthma or host-graft or graft-host disease

PS Claim 4; Page 27; 55pp; English.

XX The specification describes an immunomodulatory peptide construct.  
 CC The immunomodulatory peptide comprises a first peptide associated with  
 CC autoimmune disease, allergy or asthma, or host-versus-graft rejection  
 CC and which will bind to an antigen receptor on a set or subset of T cells,  
 CC linked to a second immune modulating peptide which will cause a directed  
 CC immune response by the set or subset of T cells to which the first  
 CC peptide is attached. Alternatively, the second peptide will bind to a  
 CC T cell receptor site on the surface of the T cell which will cause the  
 CC set or subset of T cells to initiate, but not complete, an immune  
 CC response to cause the set or subset of T cells to undergo anergy and  
 CC apoptosis. The immunomodulatory peptides are useful for eliminating a set  
 CC or subset of T cells involved in autoimmune response. They are useful for  
 CC the treatment of autoimmune disease, allergic reactions, asthma or  
 CC host-graft or graft-host rejection. The immunomodulatory peptides are  
 CC also useful for interrupting an autoimmune disease associated pathway  
 CC necessary to complete T cell activation. The present sequence represents  
 CC an immunomodulatory peptide used to treat allergic responses.

XX Sequence 42 AA;

SO Query Match 92.4%; Score 73; DB 22; Length 42;

Best Local Similarity 100.0%; Pred. No. 2.6e-06; Mismatches 0; Indels 0; Gaps 0;

OY 2 IDIFASKNFHLQKN 15  
 DB 24 IDIFASKNFHLQKN 37

## RESULT 14

AAW27369 standard; peptide; 80 AA.

AC AAW27369;

DT 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #1.

KM Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

PN WO9732600-A1.

PD 12-SEP-1997.

PF 10-MAR-1997; 97WO-JP00740.

PR 10-MAR-1996; 96JP-0080702.

PA (MEIP ) MEIJI MILK PROD CO LTD.

PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;

DR WPI, 1997-470495/43.

PT Peptide immuno:therapeutic agent to treat allergic diseases -  
PT contains multi-epitope peptide containing T cell epitope regions  
from different allergens

PS Claim 6; Page 31; 58pp; Japanese.

CC The present sequence represents a multi-epitope peptide which is used as  
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
CC or more different allergens (preferably linked via arginine or lysine  
CC dimers), where the T cell epitope regions have a positivity index  
CC greater than 100 as measured in a patient group responding to the  
CC allergen; have at least 70% reactivity with lymphocytes from patients  
CC responding to the allergen; and are not reactive with immunoglobulin E  
CC (IgE) antibodies from patients responsive to the allergen. The agent can  
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

SQ Sequence 80 AA;

Query Match 92.4%; Score 73; DB 18; Length 80;  
Best Local Similarity 100.0%; Pred. No. 5.3e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IDIFASKNPHLOKN 15

DB 31 IDIFASKNPHLOKN 44

## RESULT 15

AAW27370 standard; peptide; 105 AA.

AC AAW27370;

DT 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #2.

KM Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

PN WO9732600-A1.

PD 12-SEP-1997.

PF 10-MAR-1997; 97WO-JP00740.

PR 10-MAR-1996; 96JP-0080702.

PA (MEIP ) MEIJI MILK PROD CO LTD.

PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;

DR WPI, 1997-470495/43.

PT Peptide immuno:therapeutic agent to treat allergic diseases -  
PT contains multi-epitope peptide containing T cell epitope regions  
from different allergens

PS Claim 6; Page 31; 58pp; Japanese.

CC The present sequence represents a multi-epitope peptide which is used as  
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
CC or more different allergens (preferably linked via arginine or lysine  
CC dimers), where the T cell epitope regions have a positivity index  
CC greater than 100 as measured in a patient group responding to the  
CC allergen; have at least 70% reactivity with lymphocytes from patients  
CC responding to the allergen; and are not reactive with immunoglobulin E  
CC (IgE) antibodies from patients responsive to the allergen. The agent can  
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

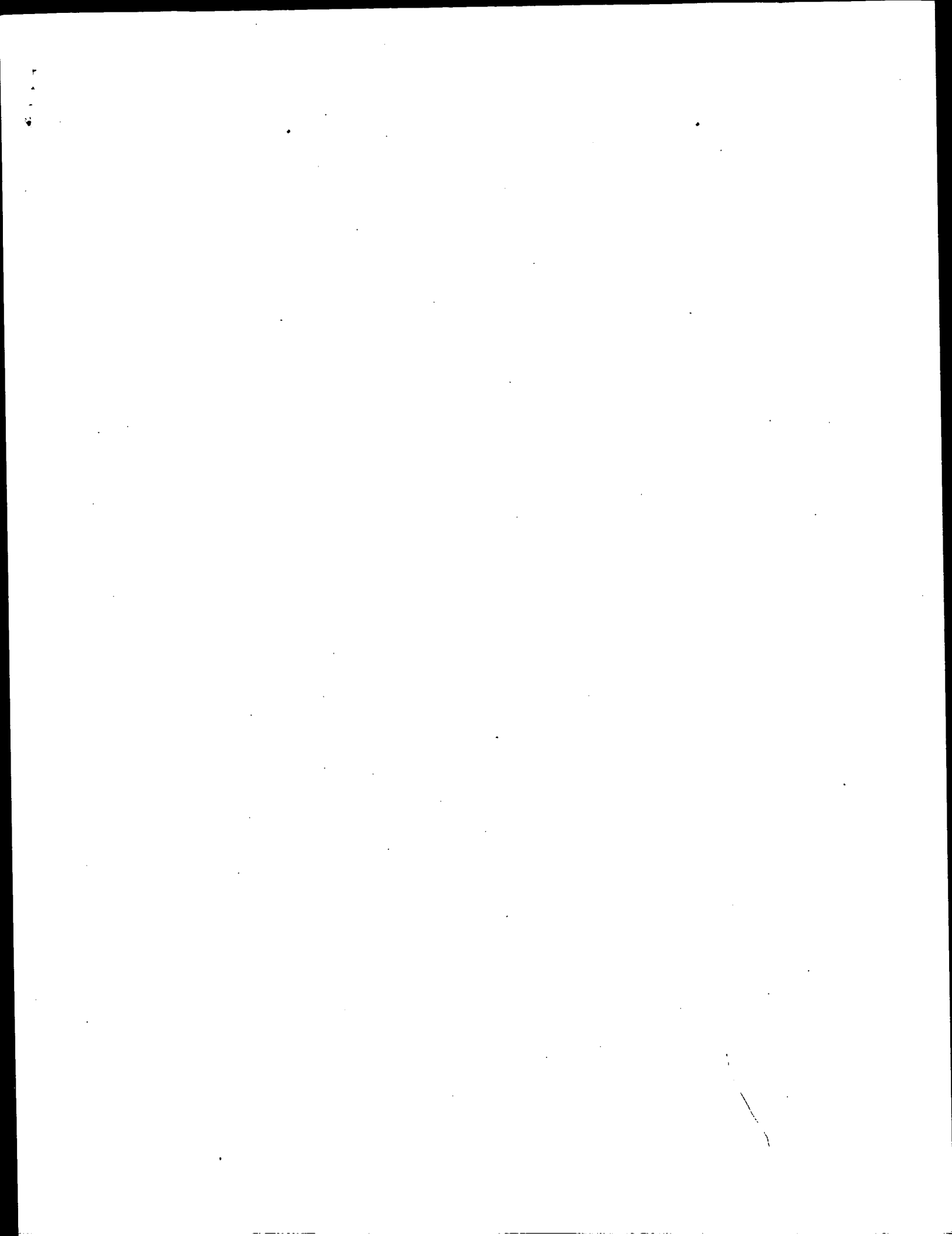
SQ Sequence 105 AA;

Query Match 92.4%; Score 73; DB 18; Length 105;  
Best Local Similarity 100.0%; Pred. No. 7.3e-06;

QY 2 IDIFASKNPHLOKN 15

DB 31 IDIFASKNPHLOKN 44

Search completed: April 20, 2003, 13:06:11  
Job time : 19.1974 secs





GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:25 ; Search time 2.92105 seconds

(without alignments)  
212.987 Million cell updates/sec

Title: US-09-142-524d-152

Perfect score: 67

Sequence: 1 LSDISLKLTKSGKIAS 15

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	67	100.0	514	MPA2_CRYJA	P43212 cryptomeria
2	43	64.2	251	ZNUC_ECOLI	P52648 escherichia
3	42	62.7	255	PECE_ECOLI	P15031 escherichia
4	38	56.7	145	VP12_AGRU	O44433 agrobacteri
5	38	56.7	316	Y428_CAREL	O76387 caenorhabdi
6	38	56.7	530	U121_HSVB	P28972 equine herp
7	37	55.2	231	YTRB_BACSU	O34392 bacillus su
8	37	55.2	232	YHCG_BACSU	P54591 bacillus su
9	37	55.2	311	Y467_MYCB	P47705 mycoplasma
10	37	55.2	454	PR11_SCHPO	O14215 schizosacch
11	37	55.2	504	PD43_MOUSE	P27773 mus musculu
12	36	53.7	229	CG18_YEAST	P38794 saccharomyc
13	36	53.7	250	MNTB_BACSU	O34338 bacillus su
14	36	53.7	365	CYSA_ECOLI	P16676 escherichia
15	36	53.7	365	CYSA_SALTY	P40860 salmonella
16	36	53.7	507	MTPS_PROST	O05474 providencia
17	36	53.7	552	Y664_HAERH	O57538 haemophilus
18	35	52.2	96	ABRB_BACSU	P08874 bacillus su
19	35	52.2	101	ILB_SHEEP	P36925 ovine aries
20	35	52.2	238	ZNUC_BUCAI	P57403 buchnera ap
21	35	52.2	264	FTSQ_STRCO	P45518 streptomyce
22	35	52.2	292	ATPG_STRMU	P95788 streptococc
23	35	52.2	339	Y467_MYCPN	P75110 mycoplasma
24	35	52.2	422	PUR2_MYCPN	P71827 mycobacteri
25	35	52.2	466	BCS1_YEAST	P32829 saccharomyc
26	35	52.2	486	COXW_YEAST	P40086 saccharomyc
27	35	52.2	560	POTA_MYCPN	P75059 mycoplasma
28	35	52.2	918	ABG5_MOUSE	O09968 mus musculu
29	35	52.2	944	PEP3_YEAST	P27801 saccharomyc
30	35	52.2	948	GLND_RHLTV	O09464 rhizobium l
31	35	52.2	948	GLND_RHLTV	O33244 rhizobium t
32	35	52.2	1186	CEAA_BACTS	O45710 bacillus th
33	35	52.2	1391	LVS2_CANAL	Q12572 candida alb

## ALIGNMENTS

RESULT 1	MPA2_CRYJA	STANDARD	PRT	514 AA.
AC	P43212			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Possible polygalacturonase precursor (RC 3.2.1.15) (PG) (Pectinase)			
DE	(Major pollen allergen Cry j 2) (Cry j II).			
OS	Cryptomeria japonica (Japanese cedar).			
OC	Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.			
OX	NCBI_TaxID=3369;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Pollen;			
RX	MEDLINE=95010777; PubMed=7926035;			
RA	Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.,			
RA	Usui M., Kurimoto M.;			
RT	"Molecular cloning of the second major allergen, Cry j II, from			
RT	Japanese cedar pollen.";			
RL	FEBS Lett. 353:124-128 (1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pollen;			
RX	MEDLINE=94271186; PubMed=8002972;			
RA	Komiyama N., Sone T., Shimizu K., Morikubo K., Kino K.;			
RT	"CDNA cloning and expression of Cry j II the second major allergen of			
RT	Japanese cedar pollen.";			
RL	Biochem. Biophys. Res. Commun. 201:1021-1028 (1994).			
RN	[3]			
RP	SEQUENCE OF 55-64.			
RC	MEDLINE=90342988; PubMed=2382797;			
RA	Sakaguchi M., Inouye S., Tanai M., Ando S., Usui M., Matunasi T.;			
RT	"Identification of the second major allergen of Japanese cedar			
RT	pollen.";			
RL	Allergy 45:309-312 (1990).			
CC	-1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-			
CC	galactosiduronic linkages in pectate and other galacturonans.			
CC	-1- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).			
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES			
CC	(POLYGALACTURONASES).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D37765; BAA07021.1; -			
DR	EMBL; D29772; BAA06172.1; -			
DR	HSSP; P26509; 1BHE.			
DR	InterPro; IPR000743; GH28.			
DR	Pfam; PF00295; Glyco_hydro_28; 1.			

PROSITE: PS00502; POLYGALACTURONASE; 1.  
 KM Hydrolyase, Glycosidase; Cell wall; Signal; Zymogen; Fruit ripening;  
 KW Amyloplase; Glycoprotein; Allergen.  
 FT SIGNAL 1 22  
 FT PROPEP 23 45  
 FT CHAIN 46 433  
 FT PROPEP 434 514  
 FT ACT SITE 278 278  
 FT CARBOHYD 460 460  
 FT CARBOHYD 472 472  
 FT CONFLICT 5 5  
 FT CONFLICT 12 12  
 FT CONFLICT 34 35  
 FT CONFLICT 37 37  
 FT CONFLICT 88 88  
 FT CONFLICT 98 98  
 FT CONFLICT 451 451  
 FT CONFLICT 454 454  
 FT CONFLICT 504 504  
 FT CONFLICT 507 507  
 SQ SEQUENCE 514 AA; 56645 MW; 62461133FAD6302 CRC64;  
 Query Match 100.0%; Score 67; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 0.00012; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0;  
 QY 1 LSDISLKTSGKIAS 15  
 DB 395 LSDISLKTSGKIAS 409  
 RESULT 2  
 ZNUC\_ECOLI STANDARD; PRT; 251 AA.  
 ID P52648; P76285;  
 AC 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE High-affinity zinc uptake system ATP-binding protein znuC.  
 GN ZNUC OR B1858 OR Z2910 OR EC82568.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / EMG2;  
 RC STRAIN=K12 / MG1655;  
 RC MEDLINE=97426617; PubMed=9278503;  
 RA Blatcher F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Coliaco-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RC MEDLINE=97251358; PubMed=9097040;  
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishino H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horichi T.,  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 40.1-50.0 min region on the linkage map."  
 RL DNA Res. 3:379-392 (1996).

[4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blatcher F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533 (2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RUMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,  
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12."  
 RL DNA Res. 8:11-22 (2001).  
 RN [6]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98343803; PubMed=9680209;  
 RA Patzer S.I., Hantke K.;  
 RT "The znuABC high-affinity zinc uptake system and its regulator Zur in  
 Escherichia coli."  
 RL Mol. Microbiol. 28:1199-1210 (1998).  
 CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY ZINC UPTAKE TRANSPORT  
 SYSTEM.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 FRAMESHIFT IN POSITION 188.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U38702; -; NOT ANNOTATED\_CDS.  
 DR EMBL; AE000280; AAC74928.1; -;  
 DR EMBL; D90828; BAA15666.1; -;  
 DR EMBL; D90829; BAA15669.1; -;  
 DR EMBL; AE005408; AAG56848.1; -;  
 DR EMBL; AP002559; BAB35991.1; -;  
 DR Ecogene; EG13132; znuC.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR Pfam; PF00005; ABC\_tran.1.  
 DR ProDom; PD000006; ABC\_transportr.1.  
 DR SMART; SM00382; AAA.1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; FALSE NEG.  
 KW ATP-binding; Transport; Zinc; Complete proteome.  
 FT NP\_BIND 37 44  
 FT CONFLICT 19 19  
 FT CONFLICT 80 80  
 FT CONFLICT 83 83  
 FT CONFLICT 135 135  
 FT CONFLICT 143 143  
 FT CONFLICT 155 155  
 FT CONFLICT 174 174  
 FT CONFLICT 176 176  
 SQ SEQUENCE 251 AA; 27867 MW; F4BF845AC3C7904C CRC64;  
 Query Match 64.2%; Score 43; DB 1; Length 251;  
 Best Local Similarity 69.2%; Pred. No. 1.4;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LSDISLKTSGK 13  
 DB 20 LSDVSLKPKGI 32

## RESULT 3

FECE\_ECOLI STANDARD; PRT; 255 AA.  
 AC P15031;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Iron(III) dicitrate transport ATP-binding protein fece.  
 GN FECE OR P4287.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=89213950; PubMed=2651410;  
 RA Steudemater H., van Hove B., Yarghi Z., Braun V.;  
 RT "Nucleotide sequences of the fecBDE genes and locations of the  
 RT proteins suggest a periplasmic-binding-protein-dependent transport  
 RT mechanism for iron(III) dicitrate in Escherichia coli.";  
 RL J. Bacteriol. 171:2626-2633(1989).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=95334362; PubMed=7610040;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blatter F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes.";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
 FOR CITRATE-DEPENDENT IRON(III). PROBABLY RESPONSIBLE FOR ENERGY  
 COUPLING TO THE TRANSPORT SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M28397; AAA23765.1; -;  
 DR EMBL; U14003; AAA97183.1; -;  
 DR EMBL; AE000499; AAC77243.1; -;  
 DR PIR; J50115; QRCM3.  
 DR EcoGene; EG10290; fece.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_Transport.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transport; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR Iron transport; Transport; Inner membrane; ATP-binding;  
 KW Complete proteome.  
 FT NP BIND 35  
 FT SEQUENCE 255 AA; 28190 MW; 89785C2A91D51F3 CRC64;

Query Match 62.7%; Score 42; DB 1; Length 255;  
 Best Local Similarity 53.3%; Pred. No. 2.2;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LSDISLKTSGK 15  
 DB 18 LSDVSLKPKGI 32

## RESULT 4

YP12\_AGRU STANDARD; PRT; 145 AA.  
 AC Q4433;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 15.8 kDa protein in plnF2.3' region (ORF2).  
 OS Agrobacterium tumefaciens.  
 OG Plasmid pTiA6.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 NCBI\_TaxID=358;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=89213933; PubMed=2708311;  
 RA Kanemoto R.H., Powell A.T., Akiyoshi D.E., Regier D.A.,  
 RA Kersteiter R.A., Nester E.W., Hawes M.C., Gordon M.P.;  
 RT "Nucleotide sequence and analysis of the plant-inducible locus plnF  
 RT from Agrobacterium tumefaciens.";  
 RL J. Bacteriol. 171:2506-2512(1989).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M19352; AAA82504.1; -;  
 DR Hypothetical 145 AA; 15829 MW; 95A3D3D8E9560AA6 CRC64;  
 SQ SEQUENCE

Query Match 56.7%; Score 38; DB 1; Length 145;  
 Best Local Similarity 66.7%; Pred. No. 6.5;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LSDISLKTSGK 12

DB 29 LSDVSLKPKGI 40

RESULT 5  
 Y248\_CAEEL STANDARD; PRT; 316 AA.  
 AC Y248\_CAEEL  
 ID 076387;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hypothetical 33.7 kDa protein C24G6.8 in chromosome V.  
 GN C24G6.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Greco T., Bradshaw H., Kepler D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0099 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF067936; AAC19215.1; -  
 DR Wormpep; C24G6.8; CB17464.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR002833; UPF0099.  
 DR Pfam; PF00627; UBA, 1.  
 DR Pfam; PF01981; UPF0099, 1.  
 DR Pfam; PD01067; UPF0099, 1.  
 DR TIGRFAMs; TIGR0283; UPF0099, 1.  
 DR Hypothetical protein.  
 SO SEQUENCE 316 AA; 33689 MW; 2084FA3D52B4EFB0 CRC64;  
 Query Match 56.7%; Score 38; DB 1; Length 316;  
 Best Local Similarity 46.7%; Pred. No. 15;  
 Matches 7; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 LSDISLKTSGKIAS 15  
 Db 206 VANSLKMTGKIAA 220  
 RESULT 6  
 UL21\_HSVB STANDARD; PRT; 530 AA.  
 AC P28872;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE Gene 40 protein.  
 GN 40.  
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 OX NCBI\_Taxid=31520;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92295566; Pubmed=1318606;  
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
 RT "The DNA sequence of equine herpesvirus-1."  
 RL Virology 189:304-316 (1992).  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL21,  
 EHV-1 40, EHV-4 UL21, AND VZV 38.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M86664; AAB02476.1; -  
 DR PIR; F36799; W2BED4.  
 DR InterPro; IPR004936; Herpes\_UL21.  
 DR Pfam; PF03252; UL21, 1.  
 SO SEQUENCE 530 AA; 57915 MW; 2E43565AB297CD18 CRC64;  
 Query Match 56.7%; Score 38; DB 1; Length 530;  
 Best Local Similarity 66.7%; Pred. No. 26;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 4 ISLKTSGKIAS 15  
 Db 126 VSLKMTGSLIVS 137  
 RESULT 7  
 YTRE\_BACSU STANDARD; PRT; 231 AA.  
 AC Q34352;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical ABC transporter ATP-binding protein ytre.

GN YTRE  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_Taxid=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; Pubmed=9384377;  
 RA Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S., Bron S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Biggeli S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabre C., Ferrari B., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Ioris B., Karamata D., Kasahara Y., Kletter-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Komagata S., Krogshøj, Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Laber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Mostel D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Pressecan E., Pujic P., Punelle B., Rapoport G., Rey M., Sadate Y.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Roche P., Scoffone P.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone P.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
 RA Tosoato V., Uchiyama S., Vandenbol M., Vannier F., Vasarotti A.,  
 RA Viani V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis."  
 RL Nature 390:249-256 (1997).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF008220; AAC00251.1; -  
 DR EMBL; Z99119; CAB15020.1; -  
 DR HSPD; Q58663; 1GSH.  
 DR Subtilisin; BG13915; YTRE.  
 DR InterPro; IPR003593; AAA\_Atpase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR NP BIND 42 ATP (POTENTIAL).  
 SO SEQUENCE 231 AA; 25460 MW; FC342312BC1A172 CRC64;  
 Query Match 55.2%; Score 37; DB 1; Length 231;  
 Best Local Similarity 50.0%; Pred. No. 16;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 LSDISLKTSGKIA 14  
 DB 25 LKOVSLVAKGEIA 38

## RESULT 8

YHCG\_BACSU STANDARD; PRT; 232 AA.  
 ID YHCG\_BACSU  
 AC P54591;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical ABC transporter ATP-binding protein yhcg.  
 GN YHCG.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97124185; PubMed=8969498;  
 RA Noback M.A., Terpestra P., Holasappel S., Venema G., Bron S.;  
 RT "A 22 kb DNA sequence in the *cspB-glpPFD* region at 75 degrees on the  
 RL Microbiology 142:3021-3026(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolyon A., Borchert S.,  
 RA Borries R., Boutsier L., Brans A., Braun M., Brigneau S.C., Bron S.,  
 RA Brouillet S., Brutschi C.V., Caldwell I.F., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Danizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Fougere D.,  
 RA Filtz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gilmour S.Y., Glaeser P., Gottlieb E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holasappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Kleaer-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maubert C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Pressan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiuchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccioni E., Takagi T., Takahashi R., Takemaru K.,  
 RA Takenuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,  
 RA Viari A., Wamboldt R., Wedler E., Weiler H., Weitzmeger T.,  
 RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RT Nature 390:249-256(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, X36983; CAA5690.1; -.

DR EMBL; Z99108; CAB12735.1; -.  
 DR Subtilist; BG1585; yhcg.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; FALSE\_NEG.  
 DR Hypothetical protein; ATP-binding; transport; Complete proteome.  
 KW NP\_BIND  
 FT NP\_BIND 35  
 FT ATP (POTENTIAL).  
 SQ SEQUENCE 232 AA; 26520 MW; 8DACDAEBC7414FF5 CRC64;

Query Match 55.2%; Score 37; DB 1; Length 232;  
 Best local similarity 46.2%; Pred. No. 16;  
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 LSDISLKTSGK 13  
 DB 18 VNDVSLTSGRI 30

## RESULT 9

Y467\_MYCGE STANDARD; PRT; 311 AA.  
 ID Y467\_MYCGE  
 AC P4705;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical ABC transporter ATP-binding protein MG467.  
 GN MG467.  
 OS Mycoplasma genitalium.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE=96026346; PubMed=7569993;  
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,  
 RA Nguyen D.T., Dougherty B.A., Sauder D.M., Phillips C.A., Merrick J.M.,  
 RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lueder T.S.,  
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 RT "The minimal gene complement of *Mycoplasma genitalium*.";  
 RT science 270:397-403(1995).  
 RN [2]  
 RP SEQUENCE OF 14-117 FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE=94075230; PubMed=8253680;  
 RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;  
 RT "A survey of the *Mycoplasma genitalium* genome by using random  
 RT sequencing.";  
 RT J. Bacteriol. 175:7918-7930(1993).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U39728; AAC72487.1; -.  
 DR EMBL; U01741; AAD10551.1; -.  
 DR TIGR; MG467;  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR Hypothetical protein; ATP-binding; transport; Complete proteome.  
 KW NP\_BIND  
 FT NP\_BIND 35  
 FT ATP (POTENTIAL).  
 SQ SEQUENCE 311 AA; 34520 MW; 8DACDAEBC7414FF5 CRC64;

```

FT  NP BIND 122 129 ATP (POTENTIAL).
SQ SEQUENCE 311 AA, 35051 MW, 678611B97C10F66B CRC64;
Query Match 55.2%; Score 37; DB 1; Length 311;
Best Local Similarity 53.8%; Pred. No. 22;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Cy 1 LSDISKLKTSNGKI 13
    |||::|||::|:|
Db 105 LSDVSLQIANGEI 117

RESULT 10
PR11 SCHPO
ID PR11 SCHPO STANDARD; PRT; 454 AA.
AC 014215;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable DNA primase small subunit (EC 2.7.7.-).
GN SPAC6B12.10C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota, Fungi, Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Frazer A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicicki G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Rger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleu V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Roeder M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerutti L., Love T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS MADE DURING DISCONTINUOUS
CC DNA REPLICATION.
CC -1- SUBUNIT: HETERODIMER OF A SMALL SUBUNIT AND A LARGE SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC PRIMASE SMALL SUBUNIT
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z98531; CAB1078.1;
CC InterPro; IPR002755; DNA_primase_S.

```

```

DR Pfam; PF01896; DNA primase_S; sm1; 1.
DR TRIGRPMs; TRIGR00335; primase_S; sm1; 1.
KW Transererase; DNA replication; DNA-directed RNA polymerase; Primosome.
FT ACT_SITE 66 66 POTENTIAL.
FT ACT_SITE 131 133 POTENTIAL.
FT ACT_SITE 133 133 POTENTIAL.
FT ACT_SITE 143 143 POTENTIAL.
FT METAL 144 144 POTENTIAL.
FT METAL 150 150 POTENTIAL.
FT METAL 153 153 POTENTIAL.
SQ SEQUENCE 454 AA; 52009 MW; 71526FFAF7C2BBFA CRC64;
Query Match 55.2%; Score 37; DB 1; Length 454;
Best Local Similarity 64.3%; Pred. No. 34;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 2 SDISFKTSGKIAS 15
Db 297 SDIVTVAAGSIAS 310
|||||
|

RESULT 11
PD33 MOUSE STANDARD; PRT; 504 AA.
ID PD33 MOUSE P27773;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein disulfide isomerase A3 precursor (BC 5.3.4.1) (Disulfide
DE isomerase ER-60) (ERP60) (58 kDa microsomal protein) (p58) (ERP57).
GN PDIA3 OR GRP58 OR ERP60 OR ERP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91237072; PubMed=2033248;
RA Hempel W.M., Defranco A.L.;
RT "Expression of phospholipase C isozymes by murine B lymphocytes.";
RN J. Immunol. 146:3713-3720(1991).
RL [2]
RP SEQUENCE OF 25-41.
RC TISSUE=Fibroblast;
RA MEDLINE=95009907; PubMed=7523108;
RX Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
RT "Separation and sequencing of familial and novel murine proteins
RT using preparative two-dimensional gel electrophoresis.";
RL Electrophoresis 15:735-745(1994).
RN [3]
RP SEQUENCE OF 25-41, AND PHOSPHORYLATION.
RC TISSUE=Fibroblast;
RA Merrick B.A., Wichter L.L., Patterson R.M., He C., Selkirk J.K.;
RT "Identification of the two isoforms of phospholipase C-alpha from
RT dividing murine fibroblasts by protein microsequencing.";
RL Biochem. Arch. 9:335-340(1993).
-1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
interchain disulfide bonds in proteins to form the native
structures.
-1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
-1- PTM: PHOSPHORYLATED.
-1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
-1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PHOSPHATIDYLINOSITOL-4,5-
BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE C-ALPHA).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```

DR EMBL: M73329; AAA39944.1; -  
 DR HSP: P07237; IMK.  
 DR SWISS-2DPAGE; P27773; MOUSE.  
 DR MGI: 95834; Gp58.  
 DR InterPro; IPR000063; ThioRed.  
 DR Pfam; PF00085; ThioRed; 2.  
 DR PRINTS; PR00421; THIOREDOXIN.  
 DR TIGRPFAMS; TIGR01126; ptd\_dcm; 2.  
 DR TIGRPFAMS; TIGR01130; ER\_PDI\_fam; 1.  
 DR PROSITE; PS00194; THIOREDOXIN; 2.  
 DR Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal;  
 KW Phosphorylation.  
 FT SIGNAL 1 24  
 FT CHAIN 25 504 PROTEIN DISULFIDE ISOMERASE A3.  
 FT DISULFID 57 60 REDOX-ACTIVE (BY SIMILARITY).  
 FT DISULFID 405 408 REDOX-ACTIVE (BY SIMILARITY).  
 FT SITE 501 504 PREVENT SECRETION FROM ER  
 (BY SIMILARITY).  
 SQ SEQUENCE 504 AA; 56621 MW; 5983B04755CEB7B4 CRC64;  
 Query Match 55.2%; Score 37; DB 1; Length 504;  
 Best Local Similarity 53.8%; Pred. No. 38;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LSDISLKTSGKI 13  
 Db 310 LSDISLKTSGI 322  
 RESULT 12  
 CG18 YEAST STANDARD; PRT; 229 AA.  
 ID CG18 YEAST STANDARD; PRT; 229 AA.  
 AC P38754;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GI/S-specific cyclin PCL5.  
 GN PCL5 OR YHR071W  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8286C / AB972;  
 RX MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaastis E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
 RA Vaudin M.;  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 VII.";  
 RT Science 265:2077-2082(1994).  
 RL -1- SIMILARITY: BELOWS TO THE CYCLIN FAMILY. MOST SIMILAR TO GI/S-  
 SPECIFIC CYCLIN PCL1 AND PCL2.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U00061; AAB68375.1; -  
 DR PIR; S46696; S46696.  
 DR SGD; S0001113; PCL5.  
 DR InterPro; IPR004366; Cyclin.  
 DR Pfam; PF00134; cyclin; 1.  
 DR SMART; SM00385; CYCLIN; 1.

DR PROSITE; PS00292; CYCLIN; FALSE NEG.  
 KW Cyclin; Cell cycle; Cell division.  
 SQ SEQUENCE 229 AA; 26467 MW; B6839ABB9D85DD49 CRC64;  
 Query Match 53.7%; Score 36; DB 1; Length 229;  
 Best Local Similarity 61.5%; Pred. No. 25;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LSDISLKTSGKI 13  
 Db 61 LSDISLKTSGKI 73  
 RESULT 13  
 MNTB BACSU STANDARD; PRT; 250 AA.  
 ID MNTB BACSU STANDARD; PRT; 250 AA.  
 AC Q34338;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Manganese transport system ATP-binding protein mntB.  
 GN MNTB.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Besterio M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Bortles R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell E., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutlan K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Colighly E.J., Grandi G.,  
 RA Guisepi G., Guy B.U., Haga K., Haeck U., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Iaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogawa K., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,  
 RA Priescan E., Pujic P., Purnelle D., Porwollik S., Prescott A.M.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sada E.,  
 RA Sato T., Scanlan E., Schleich S., Schoefer R., Scoffone P.,  
 RA Setiguchi J., Sekowaka A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,  
 RA Toato V., Uchiyama S., Vandenbol M., Vannier F., Vasseroiti A.,  
 RA Viati A., Wambut R., Wedler E., Wedler H., Welzenegger T.,  
 RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasunoto K., Yeta K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RT Nature 390:249-256(1997).  
 RL [3]  
 RP POSSIBLE FUNCTION.  
 RX MEDLINE=20223631; PubMed=10760146;  
 RA Que O., Helmann U.D.;  
 RT "Manganese homeostasis in Bacillus subtilis is regulated by MntR, a

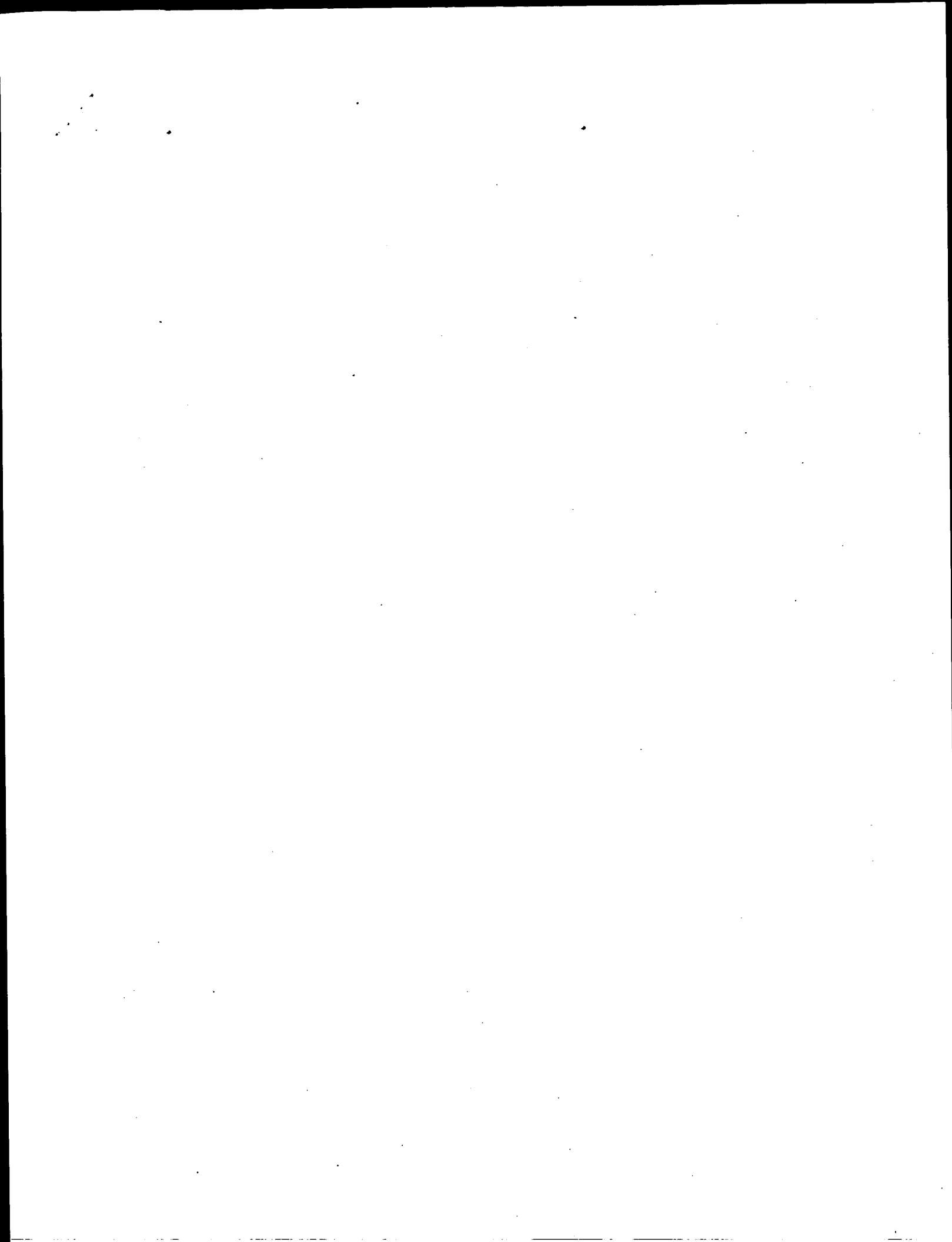
RT bifunctional regulator related to the diphtheria toxin repressor  
 RT family of proteins."  
 RL Mol. Microbiol. 35:1454-1468(2000).  
 CC -1- FUNCTION: This protein is probably a component of a manganese  
 CC perennase, a binding protein-dependent, ATP-driven transport  
 CC system. Probably responsible for energy coupling to the transport  
 CC system.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF008220; AAC00230.1; -.  
 DR EMBL, Z99119; CAB15054.1; -.  
 DR Subtilist; BG13852; mntB.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR Pfam: PF00005; ABC\_tran.1.  
 DR ProDom: PD000006; ABC\_transportr; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR Transports; ATP-binding; Membrane; Complete proteome.  
 KW NP BIND 36 43 ATP (POTENTIAL).  
 FT SEQUENCE 250 AA; 27881 MW; 12A3867A95E2208B CRC64;  
 SQ  
 Query Match 53.7%; Score 36; DB 1; Length 250;  
 Best Local Similarity 53.8%; Pred. No. 27;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 LSDISLKTSGKI 13  
 Db 19 LSDISLQVPEGKL 31  
 RESULT 14  
 CYS\_A\_ECOLI STANDARD; PRT; 365 AA.  
 AC P16676; P77693;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sulfate transport ATP-binding protein cysA.  
 GN CysA OR B2422.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 RX NCBI TaxID=562;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA MEDLINE=90264334; PubMed=2189958;  
 RA Siro A., Hymlewicz M.M., Hulanicka D.M., Boeck A.,  
 RA "Sulfate and thiosulfate transport in Escherichia coli K-12:  
 RT nucleotide sequence and expression of the cystWAM gene cluster."  
 RL J. Bacteriol. 172:3351-3357(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / WGI655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE, FROM N.A.

RC STRAIN=K12;  
 RX MEDLINE=97349980; PubMed=9205837;  
 RA Yamamoto Y., Aiba H., Baba T., Hayaishi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivaram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horuchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features."  
 RL DNA Res. 4:91-113(1997).  
 CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY SULFATE AND THIOSULFATE  
 CC MEMBRANE TRANSPORT SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, M32101; AAA2639.1; -.  
 DR EMBL, AB000329; AAC75475.1; -.  
 DR EMBL, D90872; BAA16305.1; -.  
 DR EMBL, D90871; BAA16296.1; -.  
 DR PIR; C35402; QRECSA.  
 DR EcoGene; EG10183; cysA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR Pfam; PF00005; ABC\_tran.1.  
 DR ProDom; PD000006; ABC\_transportr; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00968; aa0106801; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW Sulfate transport; Transport; ATP-binding; Inner membrane;  
 KW Complete proteome.  
 FT NP BIND 35 42 ATP (BY SIMILARITY).  
 FT CONFLICT 136 137 QL -> HV (IN REF. 1).  
 SQ SEQUENCE 365 AA; 41059 MW; BSFCC346EDF2788 CRC64;  
 Query Match 53.7%; Score 36; DB 1; Length 365;  
 Best Local Similarity 46.7%; Pred. No. 41;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 LSDISLKTSGKIAS 15  
 Db 18 LNDISLDIPSGQWVA 32  
 RESULT 15  
 CYS\_A\_SALTY STANDARD; PRT; 365 AA.  
 AC P40860;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sulfate transport ATP-binding protein cysA.  
 GN CysA OR STM2441.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=602;  
 RX NCBI TaxID=602;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGC1412 / ATCC 700720;  
 RA MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney D., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,



RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856 (2001).  
 RN [2]  
 RP SEQUENCE OF 248-365 FROM N.A.  
 RC STRAIN=LT2;  
 RA Sivaprasad A.V., Kuczek E.S., Bawden C.S., Rogers G.E.;  
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases  
 CC -! FUNCTION: INVOLVED IN THE HIGH-AFFINITY SULFATE AND THIOSULFATE  
 CC MEMBRANE TRANSPORT SYSTEM.  
 CC -! SUBCELLULAR LOCATION: Inner membrane-associated (Potential).  
 CC -! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB008810; AAL21335.1; -;  
 DR EMBL; X59595; -; NOT\_ANNOTATED\_CDS.  
 DR Stycene; SG10518; CYSA.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transportr; 1.  
 DR TIGRFAMs; TIGR00968; 3a0106601; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KM Sulfate transport; Transport; ATP-binding; Inner membrane;  
 KW Complete proteome.  
 FT NP\_BIND 35 42 ATP (BY SIMILARITY).  
 FT CONFLICT 298 303 EASPKG -> SOPEs (IN REF. 2).  
 SQ SEQUENCE 365 AA; 41035 MW; 016EAC2AF2B1C4C CRC64;  
 QY 1 LSDISLKLTKGKIAS 15  
 Db 18 LNDISLDIPSGQMVN 32  
 Query Match 53.7%; Score 36; DB 1; Length 365;  
 Best Local Similarity 46.7%; Fred. No. 41;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Search completed: April 20, 2003, 13:07:47  
 Job time : 4.92105 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524d-28

Perfect score: 81

Sequence: 1 RPLMTIFSGNMNIKL 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	20	15	AA45548
2	81	100.0	20	16	AA45548
3	81	100.0	20	16	AA45548
4	81	100.0	20	16	AA45548
5	81	100.0	20	16	AA45548
6	81	100.0	20	16	AA45548
7	81	100.0	20	16	AA45548
8	81	100.0	20	16	AA45548
9	81	100.0	20	16	AA45548
10	81	100.0	20	16	AA45548

11	81	100.0	374	15	AA60166
12	81	100.0	374	16	AA82490
13	81	100.0	374	20	AA45665
14	81	100.0	374	20	AA45665
15	62	76.5	354	17	AA42129
16	62	76.5	354	17	AA42129
17	62	76.5	354	17	AA42129
18	62	76.5	354	17	AA42129
19	62	76.5	354	17	AA42129
20	62	76.5	354	17	AA42129
21	62	76.5	354	17	AA42129
22	57	70.4	501	22	AA68273
23	57	70.4	501	22	AA68273
24	57	70.4	501	22	AA68273
25	57	70.4	501	22	AA68273
26	56	69.1	455	23	AA891703
27	56	69.1	455	23	AA891703
28	56	69.1	455	23	AA891703
29	56	69.1	455	23	AA891703
30	56	69.1	455	23	AA891703
31	56	69.1	455	23	AA891703
32	56	69.1	455	23	AA891703
33	56	69.1	455	23	AA891703
34	56	69.1	455	23	AA891703
35	56	69.1	455	23	AA891703
36	56	69.1	455	23	AA891703
37	56	69.1	455	23	AA891703
38	56	69.1	455	23	AA891703
39	56	69.1	455	23	AA891703
40	56	69.1	455	23	AA891703
41	56	69.1	455	23	AA891703
42	56	69.1	455	23	AA891703
43	56	69.1	455	23	AA891703
44	56	69.1	455	23	AA891703
45	56	69.1	455	23	AA891703

## ALIGNMENTS

RESULT 1  
AA45548  
AA45548 standard; Protein; 20 AA.  
AC AA45548;  
XX  
XX  
13-JUL-1994 (first entry)  
XX  
XX  
Cry j I pollen allergen peptide CRI-7.  
XX  
XX  
Japanese cedar; detection; allergy; treatment; diagnosis;  
XX  
XX  
T cell epitope; sensitivity.  
OS Cryptomeria japonica.  
XX  
XX  
WO9401560-A.  
XX  
XX  
20-JAN-1994.  
XX  
XX  
15-FAN-1993; 93WO-US00139.  
XX  
XX  
01-SEP-1992; 92US-0938990.  
PR 10-JUL-1992; 92WO-US05661.  
XX  
XX  
(IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
XX  
Bond UF, Garman RD, Griffith IU, Kuo M, Pollock J;  
XX  
XX  
WPI, 1994-035066/04.  
XX  
XX  
Antigens derived from Japanese cedar pollen allergen Cry j I -  
XX  
XX  
contain at least two T cell epitope(s), used to treat or diagnose  
XX  
XX  
allergy

Japanese cedar pol  
Cry j I Japanese C  
Cedar allergen 493  
Japanese cedar all  
T-cell epitope pep  
Chamaecyparis obtu  
Chamaecyparis cypres  
Jun s I. Juniperu  
Jun s I. Juniperu  
Chamaecyparis obtu  
Herbicidally activ  
Japanese cedar pol  
Residues 61-75 of  
Amino acid sequenc  
Herbicidally activ  
Herbicidally activ  
Herbicidally activ  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Part of Amb a I/An  
Ragweed allergen U  
UNC Clone 6 Amb a  
Amb a I/Antigen E  
Ragweed Pollen All  
Ragweed Amb a I.2 a  
Allergen Amb a I.2 a  
Ragweed allergen 1  
Full length Amb a  
Ragweed pollen Amb  
Ragweed pollen Amb

XX Claim 1; Fig 13; 137bp; English.  
 PS  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I (amino acids 61-80). The peptide, CJI-7,  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens.  
 CC  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 81; DB 15; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 RPLMIIFSGNNMIKL 15  
 DB 6 RPLMIIFSGNNMIKL 20  
 RESULT 2  
 ID AAR82497 standard; Protein; 20 AA.  
 AC AAR82497;  
 DT 15-APR-1996 (first entry)  
 DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-7).  
 XX Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 KM allergy; Crypomeria japonica.  
 XX Crypomeria japonica.  
 OS  
 XX  
 PN WO9527786-A1.  
 PD 19-OCT-1995.  
 PF 06-APR-1995; 95WO-US04249.  
 XX  
 PR 06-DEC-1994; 94US-0350225.  
 PR 08-APR-1994; 94US-0226248.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
 PI Shaked Z;  
 XX  
 DR WPI, 1995-366391/47.  
 XX  
 PT Modified Crypomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 XX  
 PS Disclosure; Figure 2; 60pp; English.  
 XX  
 CC Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,  
 CC modified and unmodified, are given in AAR82491-R82525. This peptide  
 CC fragment corresponds to amino acids 61-80 of the allergen mature  
 CC protein.  
 CC  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 81; DB 16; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RPLMIIFSGNNMIKL 15  
 DB 6 RPLMIIFSGNNMIKL 20  
 RESULT 3  
 ID AAM44682 standard; peptide; 30 AA.  
 AC AAM44682;  
 DT 01-MAY-1998 (first entry)  
 DE T-cell epitope peptide #1 of sugi pollen antigen.  
 XX T-cell epitope; sugi pollen antigen; sugi pollinosis.  
 KM T-cell epitope; sugi pollen antigen; sugi pollinosis.  
 XX  
 OS Synthetic.  
 OS Crypomeria japonica.  
 XX  
 PN JP10007700-A.  
 PD 13-JAN-1998.  
 XX  
 PF 24-JUN-1996; 96JP-0163287.  
 XX  
 PR 24-JUN-1996; 96JP-0163287.  
 XX  
 PA (DAIL) DAICEL CHEM IND LTD.  
 PA (MEIT) MEITI SEIKA KAISHA LTD.  
 XX  
 DR WPI, 1998-133630/13.  
 XX  
 PT T cell epitope peptide of sugi pollen antigen - useful in the  
 PT treatment of sugi pollinosis  
 XX  
 PS Claim 1; Page 1; 14pp; Japanese.  
 CC  
 CC T-cell epitope peptides AAM44682-88 and their derivatives react with  
 CC sugi pollinosis patient peripheral blood T lymphocytes. A composition  
 CC prepared by combining at least 2 of the above peptides and/or their  
 CC derivatives is used for the prevention and treatment of sugi  
 CC pollinosis.  
 CC  
 SQ Sequence 30 AA;  
 Query Match 100.0%; Score 81; DB 19; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 RPLMIIFSGNNMIKL 15  
 DB 6 RPLMIIFSGNNMIKL 20  
 RESULT 4  
 ID AAM27371 standard; peptide; 134 AA.  
 AC AAM27371;  
 DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #3.  
 XX Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KM T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 XX  
 OS Synthetic.  
 OS  
 PN WO9732600-A1.  
 XX

PD 12-SEP-1997.  
 XX  
 PF 10-MAR-1997; 97WO-JP00740.  
 XX  
 PR 10-MAR-1996; 96JP-0080702.  
 XX  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 XX  
 PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 XX  
 DR WPI; 1997-470495/43.  
 XX  
 PT Peptide immuno:therapeutic agent to treat allergic diseases.  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 XX  
 PS Claim 6; Page 32; 58pp; Japanese.  
 XX  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 CC  
 SQ Sequence 134 AA;  
 XX  
 Query Match 100.0%; Score 81; DB 18; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPLMIIFSGNNIKL 15  
 |||||  
 DB 73 RPLMIIFSGNNIKL 87  
 |||||  
 RESULT 5  
 AAR75388  
 ID AAR75388 standard; protein; 353 AA.  
 XX  
 AC AAR75388;  
 XX  
 DT 12-MAR-1996 (first entry)  
 XX  
 DE Japanese cedar pollen allergen Cryj I.  
 XX  
 KW Japanese cedar; pollen allergen; Cryj I; T-cell epitope; peptides;  
 KW prevention; treatment; cryptomeria pollinosis.  
 XX  
 OS Cryptomeria japonica.  
 OS  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 61..75 /note= "T-cell epitope peptide"  
 FT Peptide 91..105 /note= "T-cell epitope peptide"  
 FT Peptide 106..120 /note= "T-cell epitope peptide"  
 FT Peptide 146..160 /note= "T-cell epitope peptide"  
 FT Peptide 211..225 /note= "T-cell epitope peptide"  
 FT Peptide 326..340 /note= "T-cell epitope peptide"  
 FT Peptide 335..346 /note= "T-cell epitope peptide"  
 FT Peptide /note= "T-cell epitope peptide"  
 FT  
 FT JF07118295-A.  
 PN  
 XX

PD 09-MAY-1995.  
 XX  
 PF 20-OCT-1993; 93JP-0262626.  
 XX  
 PR 20-OCT-1993; 93JP-0262626.  
 XX  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 XX  
 PI WPI; 1995-203834/27.  
 XX  
 DR  
 XX  
 PT New cryptomeria pollen allergen T-cell epitope peptide - used for  
 PT prevention, treatment and investigation of Japanese cedar pollinosis  
 PT  
 PS Disclosure; Figs 1-2; 8pp; Japanese.  
 XX  
 CC AAR75388 is the Japanese cedar pollen allergen Cryj I, from which the  
 CC T-cell epitope peptides AAR81587-R89295 were derived. The peptides  
 CC can be used for the prevention and treatment of cryptomeria  
 CC pollinosis, and also for the investigation of pollinosis.  
 CC  
 SQ Sequence 353 AA;  
 XX  
 Query Match 100.0%; Score 81; DB 16; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPLMIIFSGNNIKL 15  
 |||||  
 DB 66 RPLMIIFSGNNIKL 80  
 |||||  
 RESULT 6  
 AAR81587  
 ID AAR81587 standard; protein; 353 AA.  
 XX  
 AC AAR81587;  
 XX  
 DT 24-MAY-1996 (first entry)  
 XX  
 DE Cedar pollen allergen B.  
 XX  
 KW Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KW antibody; pollinosis; therapy; immunotherapy.  
 XX  
 OS Cryptomeria japonica.  
 OS  
 XX  
 PN EP700929-A2.  
 XX  
 PD 13-MAR-1996.  
 XX  
 PF 08-SEP-1995; 95EP-0306295.  
 XX  
 PR 14-JUL-1995; 95JP-0200221.  
 PR 10-SEP-1994; 94JP-0242137.  
 PR 14-JUL-1995; 95JP-0200204.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA  
 PI Hino K, Saito S, Taniguchi Y;  
 PI  
 DR WPI; 1996-140976/15.  
 XX  
 PT New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PT used for treating cedar pollinosis  
 PT  
 PS Claim 5; Page 31-32; 36pp; English.  
 XX  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential

CC for T-cell recognition, and homologous peptides (AA081588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.

XX Sequence 353 AA;

Query Match 100.0%; Score 81; DB 17; Length 353;

Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNIKL 15  
 |||||  
 DB 66 RPLWIFSGNNIKL 80

RESULT 7  
 ID AAY25664  
 AAY25664 standard; protein; 373 AA.

AC AAY25664;

DT 30-SEP-1999 (first entry)

DE Cedar allergen 493634 Cry j IB protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

XX Cedrus sp.

XX WO9334826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 73; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a cedar (Cedrus sp.) allergen 493634 Cry j IB.

XX Sequence 373 AA;

Query Match, 100.0%; Score 81; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNIKL 15  
 |||||  
 DB 86 RPLWIFSGNNIKL 100

RESULT 8  
 ID AAY25668  
 AAY25668 standard; protein; 373 AA.

AC AAY25668;

DT 30-SEP-1999 (first entry)

DE Japanese cedar allergen 541803 Cry j I precursor protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

XX Cedrus sp.

XX WO9334826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 75; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 541803 Cry j I  
 CC precursor.

XX Sequence 373 AA;

Query Match 100.0%; Score 81; DB 20; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNIKL 15  
 |||||  
 DB 86 RPLWIFSGNNIKL 100

```

RESULT 9
ID AAR31937 standard; Protein; 374 AA.
XX
AC AAR31937;
XX
DT 03-JUN-1993 (first entry)
XX
DE Cry j I.
XX
KM Japanese cedar pollen; allergen; antigen; allergy; B cell; T cell.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Protein 22..374
FT /note= "mature Cry j I"
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PN WO9301213-A.
XX
PD 21-JAN-1993.
XX
PF 10-JUL-1992; 92WO-US05661.
XX
PR 12-JUL-1991; 91US-0729134.
PR 15-JUL-1991; 91US-0730452.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Bond JF, Griffith IU, Pollock J;
XX
DR WPI; 1993-045434/05.
DR N-PSDB; AAQ35304.
XX
PT Nucleic acid sequence encoding Cryptomeria japonica allergen -
PT for the diagnosis treatment and prevention of allergic reactions
PT to Japanese cedar pollen
XX
PS Claim 11; Page 42; 69pp; English.
XX
CC Fresh pollen and staminate cone samples were collected from a single
CC Cryptomeria japonica (Japanese cedar) tree. RNA was prep'd. and used
CC to synthesize cDNA. The cDNA was subjected to successive rounds of
CC PCR to yield a full length Cry j I clone. Cry j I or an antigenic
CC fragment of it may be used for detecting, treating and preventing an
CC allergic response to Japanese cedar pollen allergen. It is capable of
CC modifying both the B and T cell response to Cry j I and T cell response
CC to a Cry j I antigen.
XX
SQ Sequence 374 AA;
XX
Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RPLMIIFSGNMNIXL 15
DB 87 RPLMIIFSGNMNIXL 101
XX
RESULT 10
ID AAR45541 standard; Protein; 374 AA.
XX
AC AAR45541;
XX
DT 13-JUL-1994 (first entry)
XX
DE Cry j I pollen allergen.

```

```

XX
KM Japanese cedar; detection; allergy; treatment; diagnosis;
XX T cell epitope; sensitivity.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Peptide 22..374
FT /note= "mature peptide"
XX
PN WO9401560-A.
XX
PD 20-JAN-1994.
XX
PF 15-JAN-1993; 93WO-US00139.
XX
PR 01-SEP-1992; 92US-0938990.
PR 10-JUL-1992; 92WO-US05661.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;
XX
DR WPI; 1994-035066/04.
DR N-PSDB; AAQ55271.
XX
PT Antigens derived from Japanese cedar pollen allergen Cry j I -
PT contain at least two T cell epitope(s), used to treat or diagnose
PT allergy
XX
PS Disclosure; Fig 4; 137pp; English.
XX
CC The sequence is that of the Japanese cedar pollen allergen
CC Cry j I which contains at least two T cell epitopes. Peptide
CC antigens derived from it can be used for the treatment and
CC diagnosis of allergies associated with Japanese cedar pollen.
CC The peptides have enhanced therapeutic properties but reduced
CC side effects compared to naturally occurring allergens.
XX
SQ Sequence 374 AA;
XX
Query Match 100.0%; Score 81; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RPLMIIFSGNMNIXL 15
DB 87 RPLMIIFSGNMNIXL 101
XX
RESULT 11
ID AAR60166 standard; Protein; 374 AA.
XX
AC AAR60166;
XX
DT 24-MAR-1995 (first entry)
XX
DE Japanese cedar pollen antigen CryjI.
XX
KM Japanese cedar; pollen antigen; allergen; CryjI; sugi; pollinosis.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= signal_peptide
FT Protein 22..374
FT /label= mature_CryjI
XX
PN JP06197768-A.

```

XX 19-JUL-1994.  
 PD 07-JAN-1993; 93JP-0001116.  
 XX 07-JAN-1993; 93JP-0001116.  
 PR (MEIJ) MEIJI SEIKA KAISHA.  
 XX  
 PA WPI; 1994-268680/33.  
 DR N-PSDB; AAQ71601.  
 XX  
 PT Sugi (Japanese cedar) pollen antigen CryII - is useful for  
 PT diagnosis, treatment and prevention of sugi pollinosis  
 XX  
 PS Claim 2; Page 5-7; 9pp; Japanese.  
 XX  
 CC The coding sequence for the Japanese cedar ("sugi") pollen allergen  
 CC CryII was isolated from a cDNA library prepared from polyA mRNA. All  
 CC or part of the CryII protein can be used for diagnosis, treatment  
 CC and prevention of sugi pollinosis.  
 XX  
 SQ Sequence 374 AA;  
 Query Match 100.0%; Score 81; DB 15; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 3e-06; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0;  
 OY 1 RPLWIFSGNNIKL 15  
 DB 87 RPLWIFSGNNIKL 101  
 RESULT 12  
 AAR82490  
 ID AAR82490 standard; Protein; 374 AA.  
 XX  
 AC AAR82490;  
 XX  
 DT 15-APR-1996 (first entry)  
 XX  
 DE Cry j I Japanese Cedar pollen allergen.  
 XX  
 DE Cry j I Japanese cedar pollen allergen; modified; drug production;  
 KM allergy; Crytomeria japonica.  
 XX  
 OS Crytomeria japonica.  
 XX  
 PN WO9527786-A1.  
 XX  
 PD 19-OCT-1995.  
 XX  
 PF 06-APR-1995; 95WO-US04249.  
 XX  
 PR 06-DEC-1994; 94US-0350225.  
 PR 08-APR-1994; 94US-0226248.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
 PI Shaked Z;  
 XX  
 DR WPI; 1995-366391/47.  
 DR N-PSDB; AAT04248.  
 XX  
 PT Modified Crytomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 XX  
 PS Disclosure; Figure 1; 60pp; English.  
 XX  
 CC Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for

CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,  
 CC modified and unmodified, are given in AAR82491-R82525.  
 XX  
 SQ Sequence 374 AA;  
 Query Match 100.0%; Score 81; DB 16; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 3e-06; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0;  
 OY 1 RPLWIFSGNNIKL 15  
 DB 87 RPLWIFSGNNIKL 101  
 RESULT 13  
 AAY25665  
 ID AAY25665 standard; protein; 374 AA.  
 XX  
 AC AAY25665;  
 XX  
 DT 30-SEP-1999 (first entry)  
 XX  
 DE Cedar allergen 493632 Cry j IA protein fragment.  
 XX  
 DE Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 XX  
 XX WO9934826-A1.  
 XX  
 PN 15-JUL-1999.  
 XX  
 PD 11-JAN-1999; 99WO-GB00080.  
 XX  
 PF 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX  
 PI Kay AB, Larche M;  
 XX  
 DR WPI; 1999-458255/38.  
 XX  
 PT Desensitizing patients to polypeptide allergens  
 XX  
 PS Example 6; Page 73; 117pp; English.  
 XX  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, silkworm, honeybee,  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, larvae of  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a cedar (Cedrus sp.) allergen 493632 Cry j IA.



SQ Sequence 374 AA;

Query Match 100.0%; Score 81; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNIKL 15  
 |||||  
 DB 87 RPLWIFSGNNIKL 101

## RESULT 14

AAV25669  
 ID AAV25669 standard; protein; 374 AA.  
 XX  
 AC AAV25669;  
 XX  
 DT 30-SEP-1999 (first entry)  
 XX

DE Japanese cedar allergen 541802 Cry j I precursor protein fragment.  
 XX

Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mold; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX

OS Cedrus sp.  
 XX

PN W09934826-A1.  
 XX

PD 15-JUL-1999.  
 XX

PF 11-JAN-1999; 99WO-GB00080.  
 XX

PR 21-SEP-1998; 98GB-0020474.  
 XX

PR 09-JAN-1998; 98GB-0000445.  
 XX

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX

PI Kay AB, Larche M;  
 XX

DR WPI; 1999-458255/38.  
 XX

PT Desensitizing patients to polypeptide allergens  
 XX

PS Example 6; Page 75; 117pp; English.  
 XX

CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenebrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 541802 Cry j I  
 CC precursor.  
 XX

SQ Sequence 374 AA;

Query Match 100.0%; Score 81; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNIKL 15  
 |||||  
 DB 87 RPLWIFSGNNIKL 101

## RESULT 15

AAW42129  
 ID AAW42129 standard; peptide; 20 AA.  
 XX  
 AC AAW42129;  
 XX

DT 16-JUN-1998 (first entry)  
 XX

DE T-cell epitope peptide 9 from Japanese cypress pollen antigen Chaol.  
 XX

KW Japanese cypress pollen; antigen; T-cell epitope; Chaol; Chaol2;  
 KW diagnosis; allergy; spring tree pollen disease; pollinosis.  
 XX

OS Chamaecyparis obtusa.  
 XX

PN W09747648-A1.  
 XX

PD 18-DEC-1997.  
 XX

PF 12-JUN-1997; 97WO-JP02031.  
 XX

PR 14-JUN-1996; 96JP-0153527.  
 XX

PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX

PI Dairiki K, Kuno K;  
 XX

DR WPI; 1998-052242/05.  
 XX

PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
 PT Chaol and Chaol2 - used for diagnosis and treatment of spring tree  
 PT pollen disease  
 XX

PS Claim 1; Page 21; 71pp; Japanese.  
 XX

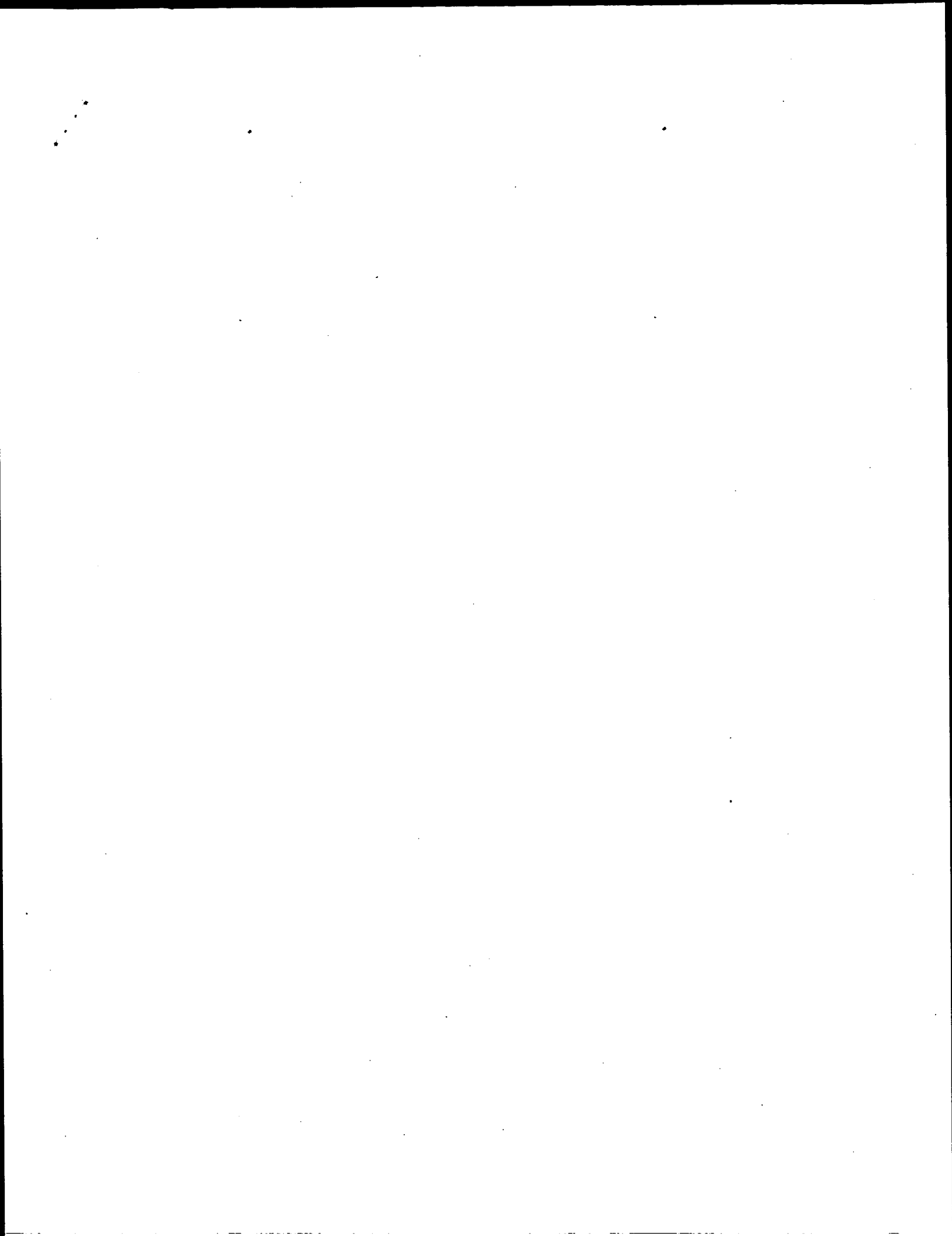
CC The present sequence represents a T-cell epitope peptide from Japanese  
 CC cypress pollen antigen Chaol. The present invention describes peptides  
 CC which correspond to the T-cell epitope sites on Japanese cypress pollen  
 CC antigens Chaol and Chaol2. The peptides can be used as a reagent for the  
 CC diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
 CC the treatment and prevention of spring tree pollen disease in which the  
 CC pollinosis involves reactivity to Japanese cypress pollen.  
 XX

SQ Sequence 20 AA;

Query Match 76.5%; Score 62; DB 19; Length 20;  
 Best Local Similarity 80.0%; Pred. No. 0.00028;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPLWIFSGNNIKL 15  
 |||||  
 DB 6 RPLWIFSGNNIKL 20

Search completed: April 20, 2003, 13:06:04  
 Job time : 19.1974 secs



concern at least two cell epitope(s), used to treat or diagnose allergy

XX Claim 1; Fig 13; 137bp; English.

PS The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I (amino acids 71-90). The peptide, Cui-8,  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens.

XX Sequence 20 AA;

Query Match 100.0%; Score 79; DB 15; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPMY 15  
 DB 1 IFSGNNNIKLKMPMY 15

RESULT 2

AA82498 AAR82498 standard; Protein; 20 AA.

AC AAR82498;

DT 15-APR-1996 (first entry)

DE Cry j I Japanese Cedar pollen allergen peptide fragment (Cui-8).

KM Cry j I; Japanese cedar pollen allergen; modified; drug production;

XX allergy; Crytpomeria japonica.

OS Crytpomeria japonica.

PN WO9527786-A1.

XX 19-OCT-1995.

PF 06-APR-1995; 95WO-US04249.

PR 06-DEC-1994; 94US-0350225.

PR 08-APR-1994; 94US-0226248.

PA (IMMU-) IMMULOGIC PHARM CORP.

PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;

PI Shaked Z;

DR WPI; 1995-366391/47.

PT Modified Crytpomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens

PS Disclosure; Figure 2; 60pp; English.

XX Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,  
 CC modified and unmodified, are given in AAR82491-R82525. This peptide  
 CC fragment corresponds to amino acids 71-90 of the allergen mature  
 CC protein.

SO Sequence 20 AA;

Query Match 100.0%; Score 79; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPMY 15  
 DB 1 IFSGNNNIKLKMPMY 15

RESULT 3

AA44682 AAM44682 standard; peptide; 30 AA.

AC AAM44682;

DT 01-MAY-1998 (first entry)

DE T-cell epitope peptide #1 of sugi pollen antigen.

KM T-cell epitope; sugi pollen antigen; sugi pollinosis.

OS Synthetic.

OS Crytpomeria japonica.

PN JP10007700-A.

PD 13-JAN-1998.

PE 24-JUN-1996; 96JP-0163287.

PR 24-JUN-1996; 96JP-0163287.

PA (DAIL) DAICEL CHEM IND LTD.

PA (MEIT) MEIJI SEIKA KAISHA LTD.

DR WPI; 1998-133630/13.

PT T cell epitope peptide of sugi pollen antigen - useful in the  
 PT treatment of sugi pollinosis

PS Claim 1; Page 1; 14pp; Japanese.

CC T-cell epitope peptides AAM44682-88 and their derivatives react with  
 CC sugi pollinosis patient peripheral blood T lymphocytes. A composition  
 CC prepared by combining at least 2 of the above peptides and/or their  
 CC derivatives is used for the prevention and treatment of sugi  
 CC pollinosis.

SO Sequence 30 AA;

Query Match 100.0%; Score 79; DB 19; Length 30;

Best Local Similarity 100.0%; Pred. No. 1.8e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPMY 15  
 DB 11 IFSGNNNIKLKMPMY 25

RESULT 4

AA27371 AAM27371 standard; peptide; 134 AA.

AC AAM27371;

DT 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #3.

KM Multi-epitope peptide; immunotherapeutic agent; allergic disease;

KM T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

PN WO9732600-A1.

PD 12-SEP-1997.  
 XX PF 10-MAR-1997; 97WO-JP00740.  
 XX PR 10-MAR-1996; 96JP-0080702.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX PI Dairiki K, Iwama A, Kino K, Kume A, Some T;  
 DR WPI; 1997-470495/43.  
 XX PT Peptide immuno-therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 XX PS Claim 6; Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 134 AA;

Query Match 100.0%; Score 79; DB 18; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IFSGNNNIKLKMPY 15  
 DB 78 IFSGNNNIKLKMPY 92

RESULT 5  
 AAR75388  
 ID AAR75388 standard; protein; 353 AA.  
 XX AAR75388;  
 AC 12-MAR-1996 (first entry)  
 XX DE Japanese cedar pollen allergen Cryj I.  
 XX DE Japanese cedar; pollen allergen; Cryj I; T-cell epitope; peptides;  
 KM prevention; treatment; cryptomeria pollinosis.  
 XX OS Cryptomeria japonica.  
 OS  
 XX Key \* Location/Qualifiers  
 FH Peptide 61..75  
 FT /note= "T-cell epitope peptide"  
 FT 91..105  
 FT /note= "T-cell epitope peptide"  
 FT 106..120  
 FT /note= "T-cell epitope peptide"  
 FT 146..160  
 FT /note= "T-cell epitope peptide"  
 FT 211..225  
 FT /note= "T-cell epitope peptide"  
 FT 326..340  
 FT /note= "T-cell epitope peptide"  
 FT 335..346  
 FT /note= "T-cell epitope peptide"  
 XX JP07118295-A.

PD 09-MAY-1995.  
 XX PF 20-OCT-1993; 93JP-0262626.  
 XX PR 20-OCT-1993; 93JP-0262626.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX PI WPI; 1995-203834/27.  
 DR  
 XX PT New cryptomeria pollen allergen T-cell epitope peptide - used for  
 PT prevention, treatment and investigation of Japanese cedar pollinosis  
 XX PS Disclosure; Figs 1-2; 8pp; Japanese.  
 CC AAR75388 is the Japanese cedar pollen allergen Cryj I, from which the  
 CC T-cell epitope peptides AAR89289-R89295 were derived. The peptides  
 CC can be used for the prevention and treatment of cryptomeria  
 CC pollinosis, and also for the investigation of pollinosis.  
 SQ Sequence 353 AA;

Query Match 100.0%; Score 79; DB 16; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IFSGNNNIKLKMPY 15  
 DB 71 IFSGNNNIKLKMPY 85

RESULT 6  
 AAR81587  
 ID AAR81587 standard; protein; 353 AA.  
 XX AAR81587;  
 AC 24-MAY-1996 (first entry)  
 XX DE Cedar pollen allergen B.  
 XX DE Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KM antibody; pollinosis; therapy; immunotherapy.  
 XX OS Cryptomeria japonica.  
 OS  
 XX EP700929-A2.  
 XX PN 13-MAR-1996.  
 XX PD 08-SEP-1995; 95EP-0306295.  
 XX PF 14-JUL-1995; 95JP-0200221.  
 PR 10-SEP-1994; 94JP-0242137.  
 PR 14-JUL-1995; 95JP-0200204.  
 XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX PI Hino K, Saito S, Taniguchi Y;  
 XX WPI; 1996-140976/15.  
 DR  
 XX PT New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 XX used for treating cedar pollinosis  
 XX PS Claim 5; Page 31-32; 36pp; English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential

CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
CC be used as immunotherapeutic agents to treat or prevent cedar  
CC pollinosis, avoiding side-effects such as anaphylaxis.

SO Sequence 353 AA;

Query Match 100.0%; Score 79; DB 17; Length 353;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNIKLKMPY 15  
Db 71 IFSGNNIKLKMPY 85

RESULT 7  
ID AAY25664  
AAY25664 standard; protein; 373 AA.

AC AAY25664;

DT 30-SEP-1999 (first entry)

DE Cedar allergen 493634 Cry j IB protein fragment.

OS Major histocompatibility complex; class II; desensitizing; human;  
KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

XX Cedrus sp.

XX WO93934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 73; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC Tenbrilio mollitor beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a cedar (Cedrus sp.) allergen 493634 Cry j IB.

XX Sequence 373 AA;

Query Match. 100.0%; Score 79; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 3.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNIKLKMPY 15  
Db 91 IFSGNNIKLKMPY 105

RESULT 8  
ID AAY25668  
AAY25668 standard; protein; 373 AA.

AC AAY25668;

DT 30-SEP-1999 (first entry)

DE Japanese cedar allergen 541803 Cry j I precursor protein fragment.

OS Major histocompatibility complex; class II; desensitizing; human;  
KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

XX Cedrus sp.

XX WO93934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 75; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC Tenbrilio mollitor beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a Japanese cedar (Cedrus sp.) allergen 541803 Cry j I  
CC precursor.

XX Sequence 373 AA;

Query Match 100.0%; Score 79; DB 20; Length 373;  
Best Local Similarity 100.0%; Pred. No. 3.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNIKLKMPY 15

Db 91 IFSGNNIKLKMPY 105

```

RESULT 9
AAR31937
ID AAR31937 standard; Protein; 374 AA.
XX
AC AAR31937;
XX
DT 03-JUN-1993 (first entry)
XX
DE Cry j I.
XX
KM Japanese cedar pollen; allergen; antigen; allergy; B cell; T cell.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Protein 22..374
FT /note= "mature Cry j I"
XX
PN WO9301213-A.
XX
PD 21-JAN-1993.
XX
PF 10-JUL-1992; 92WO-US05661.
XX
PR 12-JUL-1991; 91US-0729134.
XX
PR 15-JUL-1991; 91US-0730452.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Bond JF, Griffith J, Pollock J;
XX
DR WPI: 1993-045434/05.
XX
DR N-PSDB; AAQ5304.
XX
PT Nucleic acid sequence encoding Cryptomeria japonica allergen -
PT for the diagnosis treatment and prevention of allergic reactions
PT to Japanese cedar pollen
XX
PS Claim 11; Page 42; 69pp; English.
XX
SQ Sequence 374 AA;

Query Match 100.0%; Score 79; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPY 15
DB 92 IFSGNNNIKLKMPY 106

RESULT 10
AAR45541
ID AAR45541 standard; Protein; 374 AA.
XX
AC AAR45541;
XX
DT 13-JUL-1994 (first entry)
XX
DE Cry j I pollen allergen.

```

```

XX
KM Japanese cedar; detection; allergy; treatment; diagnosis;
XX T cell epitope; sensitivity.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Peptide 22..374
FT /note= "mature peptide"
XX
PN WO9401560-A.
XX
PD 20-JAN-1994.
XX
PF 15-JAN-1993; 93WO-US00139.
XX
PR 01-SEP-1992; 92US-0938990.
XX
PR 10-JUL-1992; 92WO-US05661.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Bond JF, Garman RD, Griffith J, Kuo M, Pollock J;
XX
DR WPI: 1994-035066/04.
XX
DR N-PSDB; AAQ55271.
XX
PT Antigens derived from Japanese cedar pollen allergen Cry j I -
PT contain at least two T cell epitope(s), used to treat or diagnose
PT allergy
XX
PS Disclosure; Fig 4; 137pp; English.
XX
SQ Sequence 374 AA;

Query Match 100.0%; Score 79; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IFSGNNNIKLKMPY 15
DB 92 IFSGNNNIKLKMPY 106

RESULT 11
AAR60166
ID AAR60166 standard; Protein; 374 AA.
XX
AC AAR60166;
XX
DT 24-MAR-1995 (first entry)
XX
DE Japanese cedar pollen antigen CryjI.
XX
DE Japanese cedar; pollen antigen; allergen; CryjI; sugi; pollinosis.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= signal_peptide
FT Protein 22..374
FT /label= mature_CryjI
XX
PN JP06197768-A.

```

```

XX 19-JUL-1994.
XX
XX 07-JAN-1993; 93JP-0001116.
XX
XX 07-JAN-1993; 93JP-0001116.
XX
XX (MEIJ) MEIJI SEIKA KAISHA.
XX
XX WPI; 1994-268680/33.
XX N-PSDB; AAQ71601.
XX
XX Sugi (Japanese cedar) pollen antigen CryjI - is useful for
XX diagnosis, treatment and prevention of sugi pollinosis
XX
XX Claim 2; Page 5-7; 9pp; Japanese.
XX
XX The coding sequence for the Japanese cedar ("sugi") pollen allergen
XX CryjI was isolated from a cDNA library prepared from polyA mRNA. All
XX or part of the CryjI protein can be used for diagnosis, treatment
XX and prevention of sugi pollinosis.
XX
XX Sequence 374 AA;
XX
XX Query Match 100.0%; Score 79; DB 15; Length 374;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 IFSGNNNIKMKMPY 15
Db 92 IFSGNNNIKMKMPY 106
XX
RESULT 12
AAR82490
ID AAR82490 standard; Protein; 374 AA.
XX
XX AAR82490;
XX
XX 15-APR-1996 (first entry)
XX
XX Cry j I Japanese Cedar pollen allergen.
XX
XX Cry j I; Japanese cedar pollen allergen; modified; drug production;
XX allergy; Crytpomeria japonica.
XX
XX Crytpomeria japonica.
XX
XX WO9527786-A1.
XX
XX 19-OCT-1995.
XX
XX 06-APR-1995; 95WO-US04249.
XX
XX 06-DEC-1994; 94US-0350225.
XX 08-APR-1994; 94US-0226248.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Chen X, Evans S, Franzen HM, Kuo M, Powers SP;
XX Shaked Z;
XX
XX WPI; 1995-366391/47.
XX N-PSDB; AAT04248.
XX
XX Modified Crytpomeria japonica (Cry j) I peptide(s) - useful for
XX treating allergy to Japanese cedar pollen allergen or
XX immunologically cross reactive allergens
XX
XX Disclosure; Figure 1; 60pp; English.
XX
XX Novel peptides of cry j I have been modified as a part of a
XX preformulation scheme to develop an optimised drug product for
XX

```

```

CC therapeutic treatment of humans suffering from allergy to Japanese
CC cedar pollen allergen or an allergen which is immunologically cross
CC reactive with Japanese cedar pollen allergen. Such modified peptides
CC possess certain characteristics which render them particularly
CC suitable for drug product formulation. Peptide fragments of Cry j I,
CC modified and unmodified, are given in AAR82491-R82525.
XX
XX Sequence 374 AA;
XX
XX Query Match 100.0%; Score 79; DB 16; Length 374;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 IFSGNNNIKMKMPY 15
Db 92 IFSGNNNIKMKMPY 106
XX
RESULT 13
AAY25665
ID AAY25665 standard; protein; 374 AA.
XX
XX AAY25665;
XX
XX 30-SEP-1999 (first entry)
XX
XX Cedar allergen 493632 Cry j IA protein fragment.
XX
XX Major histocompatibility complex; class II; desensitising; human;
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;
XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
XX cockroach; beetle; dog; horse; pig; sheep; rabbit; rat; guinea pig;
XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.
XX
XX Cedrus sp.
XX
XX WO9934826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GE00080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 73; 117pp; English.
XX
XX This invention describes a novel method of desensitising a patient to a
XX polypeptide allergen and comprises administering to the patient a peptide
XX derived from the allergen where restriction to a MHC Class II molecule
XX possessed by the patient can be demonstrated for the peptide and the
XX peptide is able to induce a late phase response in an individual who
XX possesses the MHC Class II molecule. The methods can be used for
XX desensitising patients to allergens present in e.g. grass, tree and weed
XX (including ragweed) pollens, fungi and moulds, foods, stinging insects,
XX the chironomidae (non-biting midges), spiders and mites, housefly, fruit
XX fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
XX non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
XX Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
XX sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
XX produce immunological vaccines which may be used to prevent and/or treat
XX conditions involving hypersensitivity to allergens. This sequence
XX represents a cedar (Cedrus sp.) allergen 493632 Cry j IA.
XX

```



SQ Sequence 374 AA;  
 Query Match 100.0%; Score 79; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 14

ID AAY25669 standard; protein; 374 AA.  
 AC AAY25669;

DT 30-SEP-1999 (first entry)  
 DE Japanese cedar allergen 541802 Cry j I precursor protein fragment.

Major histocompatibility complex; class II; desensitizing; human;  
 allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

OS Cedrus sp.

PN MO9934826-A1.

PD 15-JUL-1999.

PF 11-JAN-1999; 99WO-GB00080.

PR 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

PT WPI; 1999-458255/38.

PS Desensitizing patients to polypeptide allergens

Example 6; Page 75; 117pp; English.

This invention describes a novel method of desensitizing a patient to a  
 polypeptide allergen and comprises administering to the patient a peptide  
 derived from the allergen where restriction to a MHC Class II molecule  
 possessed by the patient can be demonstrated for the peptide and the  
 peptide is able to induce a late phase response in an individual who  
 possesses the MHC Class II molecule. The methods can be used for  
 desensitizing patients to allergens present in e.g. grass, tree and weed  
 (including ragweed), pollens, fungi and moulds, foods, stinging insects,  
 the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 produce immunological vaccines which may be used to prevent and/or treat  
 conditions involving hypersensitivity to allergens. This sequence  
 represents a Japanese cedar (Cedrus sp.) allergen 541802 Cry j I  
 precursor.

Sequence 374 AA;

Query Match 100.0%; Score 79; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

## RESULT 15

ID AAR45577 standard; Protein; 367 AA.  
 AC AAR45577;

DT 13-JUL-1994 (first entry)  
 DE Jun s I.

Japanese cedar; pollen allergen; allergy; treatment; diagnosis;  
 T cell epitope; sensitivity; detection.

OS Juniperus sabinoides.

Key Location/Qualifiers  
 FH Peptide 1..21 /note="signal peptide"  
 FT Peptide 22..367 /note="mature peptide"

PN MO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

PA (IMMU-) IMMUTOLOGIC PHARM CORP.

PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J;

PT WPI; 1994-035066/04.

PT N-PSDB; AA055272.

Antigen derived from Japanese cedar pollen allergen Cry j I -  
 contain at least two T cell epitope(s), used to treat or diagnose  
 allergy

PS Disclosure; Fig 16; 137pp; English.

The sequence is that of Jun s I, a homologue of the Japanese  
 cedar pollen allergen Cry j I. Antigenic peptides derived from it  
 can be used for the treatment and diagnosis of allergies associated  
 with Japanese cedar pollen.

Sequence 367 AA;

Query Match 86.1%; Score 68; DB 15; Length 367;  
 Best Local Similarity 86.7%; Pred. No. 5.2e-05;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IFSGNNNIKLKMPY 15  
 DB 92 IFSGNNNIKLKMPY 106

Search completed: April 20, 2003, 13:06:06  
 Job time : 20.1974 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds

(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524D-30

Perfect score: 79

Sequence: 1 MNTKLMPTATGK 15

Scoring table: BLOSUM62

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database: A\_Geneseq 101002.\*

Maximum Match 100%

Listing first 45 summaries

```
1: /SID27/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID27/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID27/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID27/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID27/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID27/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID27/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID27/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID27/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID27/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID27/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID27/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID27/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID27/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID27/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID27/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID27/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID27/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID27/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID27/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID27/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID27/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID27/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	20	AA45549	Cry j I pollen all
2	79	100.0	20	AA82498	Cry j I Japanese C
3	79	100.0	30	AA44682	T-cell epitope pep
4	79	100.0	134	AA27372	Multi-epitope pep
5	79	100.0	353	AA75386	Japanese cedar pol
6	79	100.0	353	AA81587	Cedar pollen aller
7	79	100.0	373	AA25664	Cedar allergen 493
8	79	100.0	374	AA25668	Japanese cedar all
9	79	100.0	374	AA31937	Cry j I. Cryptome
10	79	100.0	374	AA45541	Cry j I pollen all

11	79	100.0	374	AA60166	Japanese cedar pol
12	79	100.0	374	AA82490	Cry j I Japanese C
13	79	100.0	374	AA25665	Cedar allergen 493
14	79	100.0	374	AA25669	Japanese cedar all
15	79	88.6	357	AA45577	Japanese cedar all
16	70	88.6	370	AA45578	Japanese cedar all
17	63	79.7	105	AA27370	Multi-epitope pep
18	59	74.7	105	AA27370	T-cell epitope pep
19	59	74.7	31	AA27372	Multi-epitope pep
20	59	74.7	354	AA04344	Chamaecyparis obtu
21	59	74.7	354	AA42121	Chamaecyparis obtu
22	59	74.7	375	AA04345	Chamaecyparis obtu
23	55	69.6	20	AA45550	Cry j I pollen all
24	55	69.6	20	AA82499	Cry j I Japanese C
25	55	69.6	30	AA45579	Cry j I pollen all
26	45	57.0	24	AA45580	Cry j I pollen all
27	41	51.9	607	AA23338	Drosophila Na+ dri
28	41	51.9	773	AA23336	Drosophila Na+ dri
29	41	51.9	774	AA23340	Drosophila Na+ dri
30	41	51.9	779	AA23341	Drosophila Na+ dri
31	41	51.9	779	AA23342	Drosophila Na+ dri
32	41	51.9	1017	AA23342	Drosophila Na+ dri
33	41	51.9	1030	AA23337	Drosophila Na+ dri
34	41	51.9	1030	AA23339	Drosophila Na+ dri
35	41	51.9	1086	AA23339	Drosophila Na+ dri
36	39	49.4	97	ABP01586	Human ORFX protein
37	39	49.4	162	AA38669	Neisseria meningit
38	38	48.1	20	AA42131	T-cell epitope pep
39	38	48.1	42	AA82824	Human immune haema
40	38	48.1	320	AA85420	Lactococcus lactis
41	38	48.1	444	ABB6690	Drosophila melanog
42	38	48.1	551	AA52398	Human keratin KRT
43	38	48.1	551	AA52398	Human keratin-2 (k
44	37	46.8	74	AA20423	Human secreted pro
45	37	46.8	104	AA04575	Human polypeptide

#### ALIGNMENTS

```
RESULT 1
ID   AAR45549 standard; Protein; 20 AA.
AC   AAR45549;
XX
XX
DT   13-JUL-1994 (first entry)
DE   Cry j I pollen allergen peptide Cyt-8.
XX
XX   Japanese cedar; detection; allergy; treatment; diagnosis;
XX   T cell epitope; sensitivity.
XX
XX   Cryptomeria-japonica.
XX
XX   WO9401560-A.
XX
XX   20-JAN-1994.
XX
XX
XX   IS-JAN-1993; 93WO-US00139.
XX
XX   01-SEP-1992; 92US-0938990.
XX   PR   10-JUL-1992; 92WO-US05661.
XX
XX   (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX   Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;
XX   WPI; 1994-035066/04.
XX
XX   Antigen derived from Japanese cedar pollen allergen Cry j I -
XX   contain at least two T cell epitope(s), used to treat or diagnose
XX   allergy
```

XX Claim 1, Fig 13; 137bp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j I (amino acids 71-90). The peptide, CJI-8,  
CC can be used for the treatment and diagnosis of allergies associated  
CC with Japanese cedar pollen. It has enhanced therapeutic properties  
CC but reduced side effects compared to naturally occurring allergens.

XX Sequence 20 AA;

Query Match 100.0%; Score 79; DB 15; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.3e-08; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNIKLKMPYIAGYK 15  
DB 6 MNIKLKMPYIAGYK 20

RESULT 2  
ID AAR82498 standard; Protein; 20 AA.

XX AAR82498;

DT 15-APR-1996 (first entry)

DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-8).

KW Cry j I; Japanese cedar pollen allergen; modified; drug production;  
XX allergy; Cryptomeria japonica.

OS Cryptomeria japonica.

PN WO9527786-A1.

PD 19-OCT-1995.

PF 06-APR-1995; 95WO-US04249.

PR 06-DEC-1994; 94US-0350225.

XX 08-APR-1994; 94US-0226248.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;

XX Shaked Z;

PT Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for  
PT treating allergy to Japanese cedar pollen allergen or  
PT immunologically cross reactive allergens

PS Disclosure; Figure 2; 60pp; English.

XX Novel peptides of cry j I have been modified as a part of a  
CC preformulation scheme to develop an optimised drug product for  
CC therapeutic treatment of humans suffering from allergy to Japanese  
CC cedar pollen allergen or an allergen which is immunologically cross  
CC reactive with Japanese cedar pollen allergen. Such modified peptides  
CC possess certain characteristics which render them particularly  
CC suitable for drug product formulation. Peptide fragments of Cry j I,  
CC modified and unmodified, are given in AAR82491-R82525. This peptide  
CC fragment corresponds to amino acids 71-90 of the allergen mature  
CC protein.

SQ Sequence 20 AA;

Query Match 100.0%; Score 79; DB 16; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.3e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNIKLKMPYIAGYK 15  
DB 6 MNIKLKMPYIAGYK 20

RESULT 3  
ID AAW44682 standard; peptide; 30 AA.

XX AAW44682;

DT 01-MAY-1998 (first entry)

DE T-cell epitope peptide #1 of sugi pollen antigen.

KW T-cell epitope; sugi pollen antigen; sugi pollinosis.

OS Synthetic.

XX Cryptomeria japonica.

PN JP10007700-A.

PD 13-JAN-1998.

PF 24-JUN-1996; 96UP-0163287.

PR 24-JUN-1996; 96UP-0163287.

PA (DAIL ) DAICEL CHEM IND LTD.

XX (MEIJ ) MEIJI SEIKA KAISHA LTD.

DR WPI, 1998-133630/13.

PT T cell epitope peptide of sugi pollen antigen - useful in the  
PT treatment of sugi pollinosis

PS Claim 1; Page 1; 14pp; Japanese.

CC T-cell epitope peptides AAW44682-88 and their derivatives react with  
CC sugi pollinosis patient peripheral blood T lymphocytes. A composition  
CC prepared by combining at least 2 of the above peptides and/or their  
CC derivatives is used for the prevention and treatment of sugi  
CC pollinosis.

SQ Sequence 30 AA;

Query Match 100.0%; Score 79; DB 19; Length 30;  
Best Local Similarity 100.0%; Pred. No. 6.9e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNIKLKMPYIAGYK 15  
DB 16 MNIKLKMPYIAGYK 30

RESULT 4

ID AAW27371 standard; peptide; 134 AA.

XX AAW27371;

DT 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #3.

KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;

XX T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

PN WO9732600-A1.



CC for T-cell recognition, and homologous peptides (AAR1588-96) can  
CC be used as immunotherapeutic agents to treat or prevent cedar  
CC pollinosis, avoiding side-effects such as anaphylaxis.

XX Sequence 353 AA;

Query Match 100.0%; Score 79; DB 17; Length 353;

Best Local Similarity 100.0%; Pred. No. 1.2e-06; Mismatches 0; Gaps 0;

Matches 15; Conservative 0; Indels 0; Gaps 0;

OY 1 MNIKLKMPTIAGYK 15  
DB 76 MNIKLKMPTIAGYK 90

RESULT 7  
ID AAY25664 standard; protein; 373 AA.

XX AAY25664;

DT 30-SEP-1999 (first entry)

DE Cedar allergen 493634 Cry j IB protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
XX cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

OS Cedrus sp.

PN WO9934826-A1.

PD 15-JUL-1999.

PF 11-JAN-1999; 99WO-GB00080.

PR 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

DR WPI; 1999-458255/38.

PT Desensitizing patients to polypeptide allergens

PS Example 6; Page 73; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a cedar (Cedrus sp.) allergen 493634 Cry j IB.

XX Sequence 373 AA;

Query Match 100.0%; Score 79; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 1.3e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNIKLKMPTIAGYK 15

DB 96 MNIKLKMPTIAGYK 110

RESULT 8

ID AAY25668 standard; protein; 373 AA.

XX AAY25668;

DT 30-SEP-1999 (first entry)

DE Japanese cedar allergen 541803 Cry j I precursor protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
XX cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

OS Cedrus sp.

PN WO9934826-A1.

PD 15-JUL-1999.

PF 11-JAN-1999; 99WO-GB00080.

PR 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

DR WPI; 1999-458255/38.

PT Desensitizing patients to polypeptide allergens

PS Example 6; Page 75; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a Japanese cedar (Cedrus sp.) allergen 541803 Cry j I  
CC precursor.

XX Sequence 373 AA;

Query Match 100.0%; Score 79; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 1.3e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNIKLKMPTIAGYK 15

DB 96 MNIKLKMPTIAGYK 110

RESULT 9  
 AAR31937 standard; Protein; 374 AA.  
 ID AAR31937  
 AC AAR31937;  
 XX  
 XX  
 DT 03-JUN-1993 (first entry)  
 DE Cry j I.  
 XX  
 KW Japanese cedar pollen; allergen; antigen; allergy; B cell; T cell.  
 OS Cryptomeria japonica.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /note= "signal peptide"  
 FT Protein 22..374 /note= "mature Cry j I"  
 FT  
 XX  
 XX  
 PA WO9301213-A.  
 XX  
 XX  
 PD 21-JAN-1993.  
 XX  
 PF 10-JUL-1992; 92WO-US05661.  
 XX  
 PR 12-JUL-1991; 91US-0729134.  
 PR 15-JUL-1991; 91US-0730452.  
 XX  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Bond JF, Griffith J, Pollock J;  
 XX  
 DR WPI; 1993-045434/05.  
 DR N-PSDB; AAQ35304.  
 XX  
 PT Nucleic acid sequence encoding Cryptomeria japonica allergen -  
 PT for the diagnosis treatment and prevention of allergic reactions  
 PT to Japanese cedar pollen  
 XX  
 PS Claim 11; Page 42; 69pp; English.  
 XX  
 CC Fresh pollen and staminate cone samples were collected from a single  
 CC Cryptomeria japonica (Japanese cedar) tree. RNA was prep'd. and used  
 CC to synthesize cDNA. The cDNA was subjected to successive rounds of  
 CC PCR to yield a full length Cry j I clone. Cry j I or an antigenic  
 CC fragment of it may be used for detecting, treating and preventing an  
 CC allergic response to Japanese cedar pollen allergen. It is capable of  
 CC modifying both the B and T cell response to Cry j I and T cell response  
 CC to a Cry j I antigen.  
 CC  
 SO Sequence 374 AA;  
 Query Match 100.0%; Score 79; DB 14; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNIKLKMPTIAGYK 15  
 DB 97 MNIKLKMPTIAGYK 111  
 DE  
 RESULT 10  
 AAR45541  
 ID AAR45541 standard; Protein; 374 AA.  
 AC AAR45541;  
 XX  
 XX  
 DT 13-JUL-1994 (first entry)  
 DE Cry j I pollen allergen.

XX  
 KW Japanese cedar; detection; allergy; treatment; diagnosis;  
 KM T cell epitope; sensitivity.  
 XX  
 XX  
 OS Cryptomeria japonica.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /note= "signal peptide"  
 FT Peptide 22..374 /note= "mature peptide"  
 FT  
 XX  
 XX  
 PA WO9401560-A.  
 XX  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US00139.  
 XX  
 PR 01-SEP-1992; 92US-0938990.  
 PR 10-JUL-1992; 92WO-US05661.  
 XX  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 XX  
 PI Bond JF, Garman RD, Griffith J, Kuo M, Pollock J;  
 XX  
 XX  
 DR WPI; 1994-035066/04.  
 DR N-PSDB; AAQ55271.  
 XX  
 PT Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 XX  
 PS Disclosure; Fig 4; 137pp; English.  
 XX  
 CC The sequence is that of the Japanese cedar pollen allergen  
 CC Cry j I which contains at least two T cell epitopes. Peptide  
 CC antigens derived from it can be used for the treatment and  
 CC diagnosis of allergies associated with Japanese cedar pollen.  
 CC The peptides have enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 CC  
 SO Sequence 374 AA;  
 Query Match 100.0%; Score 79; DB 15; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNIKLKMPTIAGYK 15  
 DB 97 MNIKLKMPTIAGYK 111  
 DE  
 RESULT 11  
 AAR60166  
 ID AAR60166 standard; Protein; 374 AA.  
 AC AAR60166;  
 XX  
 XX  
 DT 24-MAR-1995 (first entry)  
 DE Japanese cedar pollen antigen CryjI.  
 XX  
 XX  
 KW Japanese cedar; pollen antigen; allergen; CryjI; sugi; pollinosis.  
 OS Cryptomeria japonica.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /label= "signal\_peptide"  
 FT Protein 22..374 /label= "mature\_CryjI"  
 FT  
 XX  
 PA JP06197768-A.

XX 19-JUL-1994.  
 PD 07-JAN-1993; 93JP-0001116.  
 XX 07-JAN-1993; 93JP-0001116.  
 PR (MEIJ) MEIJI SEIKA KAISHA.  
 XX N-PSDB; AAQ71601.  
 DR N-PSDB; AAQ71601.  
 XX Sugi (Japanese cedar) pollen antigen Cryj1 - is useful for  
 PT diagnosis, treatment and prevention of sugi pollinosis  
 PS Claim 2; Page 5-7; 9pp; Japanese.  
 XX The coding sequence for the Japanese cedar ("sugi") pollen allergen  
 CC Cryj1 was isolated from a cDNA library prepared from polyA mRNA. All  
 CC or part of the Cryj1 protein can be used for diagnosis, treatment  
 CC and prevention of sugi pollinosis.  
 XX Sequence 374 AA;  
 SQ  
 Query Match 100.0%; Score 79; DB 15; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNIKMKMPYIAGYK 15  
 Db 97 MNIKMKMPYIAGYK 111  
 RESULT 12  
 AAR82490  
 ID AAR82490 standard; Protein; 374 AA.  
 AC AAR82490;  
 XX 15-APR-1996 (first entry)  
 DT Cry j I Japanese cedar pollen allergen.  
 XX Cry j I Japanese cedar pollen allergen; modified; drug production;  
 DE Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 XX Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 KM Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 OS Crypomeria japonica.  
 XX Crypomeria japonica.  
 PN MO9527786-A1.  
 PD 19-OCT-1995.  
 PF 06-APR-1995; 95WO-US04249.  
 PR 06-DEC-1994; 94US-0350225.  
 XX 08-APR-1994; 94US-0226248.  
 XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PA Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
 PI Shaked Z;  
 XX WPI; 1995-366391/47.  
 DR N-PSDB; AAT04248.  
 XX Modified Crypomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 XX Disclosure; Figure 1; 60pp; English.  
 PS Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for

CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,  
 CC modified and unmodified, are given in AAR82491-R82525.  
 XX  
 SQ Sequence 374 AA;  
 Query Match 100.0%; Score 79; DB 16; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNIKMKMPYIAGYK 15  
 Db 97 MNIKMKMPYIAGYK 111  
 RESULT 13  
 AAY25665  
 ID AAY25665 standard; protein; 374 AA.  
 AC AAY25665;  
 XX 30-SEP-1999 (first entry)  
 DT Cedar allergen 493632 Cry j IA protein fragment.  
 XX Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mold; food; insect; sting;  
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX Cedrus sp.  
 OS  
 XX WO9934826-A1.  
 PN 15-JUL-1999.  
 PD 11-JAN-1999; 99WO-GB00080.  
 PF 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PA Kay AB, Larche M;  
 PI WPI; 1999-458255/38.  
 DR Desensitizing patients to polypeptide allergens  
 XX Example 6; Page 73; 117pp; English.  
 PS This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The method can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a cedar (Cedrus sp.) allergen 493632 Cry j IA.



SQ Sequence 374 AA;  
 Query Match 100.0%; Score 79; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPMYIAGYK 15  
 |||||  
 DB 97 MNIKLKMPMYIAGYK 111

RESULT 14  
 AAY25669  
 ID AAY25669 standard; protein; 374 AA.  
 AC AAY25669;  
 XX  
 XX 30-SEP-1999 (first entry)  
 DT  
 XX Japanese cedar allergen 541802 Cry j I precursor protein fragment.  
 DE  
 XX Major histocompatibility complex; class II; desensitizing; human;  
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;  
 KM chironomidae; spider; mice; housefly; fruit fly; sheep blow fly; honeybee;  
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 XX  
 PN WO9934826-A1.  
 XX  
 PD 15-JUL-1999.  
 XX  
 PF 11-JAN-1999; 99WO-GB00080.  
 XX  
 PR 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX  
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PA  
 XX  
 PI Kay AB, Larche M;  
 XX  
 DR WPI; 1999-458255/38.  
 XX  
 PT Desensitizing patients to polypeptide allergens  
 XX  
 PS Example 6; Page 75; 117pp; English.  
 XX  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrilio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 541802 Cry j I  
 CC precursor.  
 CC  
 SQ Sequence 374 AA;  
 XX  
 XX Query Match 100.0%; Score 79; DB 20; Length 374;  
 XX Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPMYIAGYK 15  
 |||||  
 DB 97 MNIKLKMPMYIAGYK 111

RESULT 15  
 AAR45577  
 ID AAR45577 standard; Protein; 367 AA.  
 AC AAR45577;  
 XX  
 XX 13-JUL-1994 (first entry)  
 DT  
 XX Jun s I.  
 DE  
 XX Japanese cedar; pollen allergen; allergy; treatment; diagnosis;  
 KM T cell epitope; sensitivity; detection.  
 KM  
 XX Juniperus sabinoides.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "signal peptide"  
 FT Peptide 22..367  
 FT /note= "mature peptide"  
 PN WO9401560-A.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US00139.  
 XX  
 PR 01-SEP-1992; 92US-0938990.  
 PR 10-JUL-1992; 92WO-US05661.  
 XX  
 XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PA  
 XX Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J;  
 PI WPI; 1994-035066/04.  
 XX  
 DR N-PSDB; AAQ55272.  
 XX  
 XX Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 PS Disclosure; Fig 16; 137pp; English.  
 XX  
 CC The sequence is that of Jun s I, a homologue of the Japanese  
 CC cedar pollen allergen Cry j I. Antigenic peptides derived from it  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen.  
 CC  
 SQ Sequence 367 AA;  
 XX  
 XX Query Match 88.6%; Score 70; DB 15; Length 367;  
 XX Best Local Similarity 80.0%; Pred. No. 6.4e-05;  
 XX Matches 12; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKLKMPMYIAGYK 15  
 |||||  
 DB 97 MNIKLKMPMYIAGYK 111

Search completed: April 20, 2003, 13:06:06  
 Job time : 18.1974 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds

(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524D-31

Perfect score: 83  
Sequence: 1 KMPMYTAGYKTFDGR 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83	100.0	20	15	AA845550
2	83	100.0	20	16	AA845550
3	83	100.0	30	15	AA845550
4	83	100.0	105	18	AA845550
5	83	100.0	134	18	AA845550
6	83	100.0	353	16	AA845550
7	83	100.0	353	17	AA845550
8	83	100.0	373	20	AA845550
9	83	100.0	373	20	AA845550
10	83	100.0	374	14	AA845550

11	83	100.0	374	15	AA845551
12	83	100.0	374	15	AA845551
13	83	100.0	374	15	AA845551
14	83	100.0	374	20	AA845551
15	83	100.0	374	20	AA845551
16	83	100.0	374	20	AA845551
17	83	100.0	374	20	AA845551
18	83	100.0	374	20	AA845551
19	83	100.0	374	20	AA845551
20	83	100.0	374	20	AA845551
21	83	100.0	374	20	AA845551
22	83	100.0	374	20	AA845551
23	83	100.0	374	20	AA845551
24	83	100.0	374	20	AA845551
25	83	100.0	374	20	AA845551
26	83	100.0	374	20	AA845551
27	83	100.0	374	20	AA845551
28	83	100.0	374	20	AA845551
29	83	100.0	374	20	AA845551
30	83	100.0	374	20	AA845551
31	83	100.0	374	20	AA845551
32	83	100.0	374	20	AA845551
33	83	100.0	374	20	AA845551
34	83	100.0	374	20	AA845551
35	83	100.0	374	20	AA845551
36	83	100.0	374	20	AA845551
37	83	100.0	374	20	AA845551
38	83	100.0	374	20	AA845551
39	83	100.0	374	20	AA845551
40	83	100.0	374	20	AA845551
41	83	100.0	374	20	AA845551
42	83	100.0	374	20	AA845551
43	83	100.0	374	20	AA845551
44	83	100.0	374	20	AA845551
45	83	100.0	374	20	AA845551

## ALIGNMENTS

RESULT 1  
ID AA845550 standard; Protein; 20 AA.  
AC AA845550;  
XX  
DT 13-JUN-1994 (first entry)  
XX  
DE Cry j I pollen allergen peptide CUI-9.  
XX  
KW Japanese cedar; detection; allergy; treatment; diagnosis;  
XX T cell epitope; sensitivity.  
XX  
OS Cryptomeria japonica.  
XX  
PN W09401560-A.  
XX  
PD 20-JAN-1994.  
XX  
PF 15-JAN-1993; 93WO-US00139.  
XX  
PR 01-SEP-1992; 92US-0938990.  
XX 10-JUN-1992; 92WO-US05661.  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;  
XX WPI; 1994-035066/04.  
XX  
XX Antigen derived from Japanese cedar pollen allergen Cry j I -  
PT contain at least two T cell epitope(s), used to treat or diagnose  
PT allergy

XX PS Claim 1; Fig 13; 137pp; English.  
 XX CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I (amino acids 81-100). The peptide, CJI-9,  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens.  
 XX SQ Sequence 20 AA;  
 Query Match 100.0%; Score 83; DB 15; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KMPMYIAGYKTFDGR 15  
 DB 1 KMPMYIAGYKTFDGR 15  
 RESULT 2  
 AAR82499 ID AAR82499 standard; Protein; 20 AA.  
 XX AC AAR82499;  
 XX DT 15-APR-1996 (first entry)  
 XX DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-9).  
 XX KM Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 XX KW allergy; Cryptomeria japonica.  
 XX OS Cryptomeria japonica.  
 XX PN WO9527786-A1.  
 XX PD 19-OCT-1995.  
 XX PF 06-APR-1995; 95WO-US04249.  
 XX PR 06-DEC-1994; 94US-0350225.  
 XX PR 08-APR-1994; 94US-0226248.  
 XX PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
 XX PI Shaked Z;  
 XX DR WPI, 1995-366391/47.  
 XX PT Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 XX PS Disclosure; Figure 2; 60pp; English.  
 XX CC Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,  
 CC modified and unmodified, are given in AAR82491-R82525. This peptide  
 CC fragment corresponds to amino acids 81-100 of the allergen mature  
 CC protein.  
 XX SQ Sequence 20 AA;  
 Query Match 100.0%; Score 83; DB 16; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMPMYIAGYKTFDGR 15  
 DB 1 KMPMYIAGYKTFDGR 15  
 RESULT 3  
 AAR45579 ID AAR45579 standard; Protein; 30 AA.  
 XX AC AAR45579;  
 XX DT 13-JUL-1994 (first entry)  
 XX DE Cry j I pollen allergen peptide CJI-41.  
 XX KM Japanese cedar; detection; allergy; treatment; diagnosis;  
 XX KW T cell epitope; sensitivity.  
 XX OS Cryptomeria japonica.  
 XX PN WO9401560-A.  
 XX PD 20-JAN-1994.  
 XX PF 15-JAN-1993; 93WO-US00139.  
 XX PR 01-SEP-1992; 92US-0938890.  
 XX PR 10-JUL-1992; 92WO-US05661.  
 XX PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J;  
 XX DR WPI, 1994-035066/04.  
 XX PT Antigens derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX PS Claim 76; Fig 18; 137pp; English.  
 XX CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I. The peptide, CJI-41, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 XX SQ Sequence 30 AA;  
 Query Match 100.0%; Score 83; DB 15; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KMPMYIAGYKTFDGR 15  
 DB 1 KMPMYIAGYKTFDGR 15  
 RESULT 4  
 AAM27370 ID AAM27370 standard; peptide; 105 AA.  
 XX AC AAM27370;  
 XX DT 24-MAR-1998 (first entry)  
 XX DE Multi-epitope peptide used as immunotherapeutic agent #2.  
 XX KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 XX KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 XX OS Synthetic.

XX MO9732600-A1.  
 PN 12-SEP-1997.  
 PD 10-MAR-1997; 97WO-JP00740.  
 PF 10-MAR-1996; 96JP-0080702.  
 PR (MEIP ) MEIJI MILK PROD CO LTD.  
 PA Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 PI WPI; 1997-470495/43.  
 DR Peptide immuno:therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 from different allergens  
 PS Claim 6; Page 31; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 CC  
 SQ Sequence 105 AA;  
 Query Match 100.0%; Score 83; DB 18; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KMPMYIAGYKTFDGR 15  
 DB 75 KMPMYIAGYKTFDGR 89  
 RESULT 5  
 AAW27371  
 ID AAW27371 standard; peptide; 134 AA.  
 AC AAW27371;  
 XX 24-MAR-1998 (first entry)  
 DT Multi-epitope peptide used as immunotherapeutic agent #3.  
 DE Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 XX Synthetic.  
 OS WO9732600-A1.  
 PN 12-SEP-1997.  
 PD 10-MAR-1997; 97WO-JP00740.  
 PF 10-MAR-1996; 96JP-0080702.  
 PR (MEIP ) MEIJI MILK PROD CO LTD.  
 PA Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 PI WPI; 1997-470495/43.  
 DR Peptide immuno:therapeutic agent to treat allergic diseases -

PT contains multi-epitope peptide containing T cell epitope regions  
 from different allergens  
 PS Claim 6; Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 CC  
 SQ Sequence 134 AA;  
 Query Match 100.0%; Score 83; DB 18; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KMPMYIAGYKTFDGR 15  
 DB 88 KMPMYIAGYKTFDGR 102  
 RESULT 6  
 AAR75388  
 ID AAR75388 standard; protein; 353 AA.  
 AC AAR75388;  
 XX 12-MAR-1996 (first entry)  
 DT Japanese cedar pollen allergen Cry I.  
 DE Japanese cedar pollen allergen Cry I.  
 KW Japanese cedar; pollen allergen; Cry I; T-cell epitope; peptides;  
 KW prevention; treatment; cryptomeria pollinosis.  
 XX Cryptomeria japonica.  
 OS  
 XX Key Location/Qualifiers  
 FH Peptide 61..75  
 FT /note= "T-cell epitope peptide"  
 FT 91..105  
 FT /note= "T-cell epitope peptide"  
 FT 106..120  
 FT /note= "T-cell epitope peptide"  
 FT 146..160  
 FT /note= "T-cell epitope peptide"  
 FT 211..225  
 FT /note= "T-cell epitope peptide"  
 FT 326..340  
 FT /note= "T-cell epitope peptide"  
 FT 335..346  
 FT /note= "T-cell epitope peptide"  
 JP07118295-A.  
 PN 09-MAY-1995.  
 PD 20-OCT-1993; 93JP-0262626.  
 PF 20-OCT-1993; 93JP-0262626.  
 PR (MEIP ) MEIJI MILK PROD CO LTD.  
 PA WPI; 1995-203834/27.  
 XX New cryptomeria pollen allergen T-cell epitope peptide - used for  
 PT prevention, treatment and investigation of Japanese cedar pollinosis

PS Disclosure; Figs 1-2; 8pp; Japanese.

CC AAR5388 is the Japanese cedar pollen allergen Cry I, from which the  
CC T-cell epitope peptides AAR8289-R89295 were derived. The peptides  
CC can be used for the prevention and treatment of cryptomera  
CC pollinosis, and also for the investigation of pollinosis.

XX Sequence 353 AA;

SO Query Match 100.0%; Score 83; DB 16; Length 353;

Best Local Similarity 100.0%; Pred. No. 1.2e-06; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMPMYIAGYKTFDGR 15  
DB 81 KMPMYIAGYKTFDGR 95

RESULT 7

AAR81587 standard; Protein; 353 AA.

XX AAR81587;

XX 24-MAY-1996 (first entry)

XX Cedar pollen allergen B.

XX Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;

XX antibody; pollinosis; therapy; immunotherapy.

XX Cryptomera japonica.

XX EP700929-A2.

XX 13-MAR-1996.

XX 08-SEP-1995; 95EP-0306295.

XX 14-JUL-1995; 95JP-0200221.

XX 10-SEP-1994; 94JP-0242137.

XX 14-JUL-1995; 95JP-0200204.

XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

XX Hino K, Saito S, Taniguchi Y;

XX WPI; 1996-140976/15.

XX New peptide(s) derived from cedar pollen allergens - activate

XX allergen-specific T-cells, but not allergen-specific IgE antibodies,

XX used for treating cedar pollinosis

XX Claim 5; Page 31-32; 36pp; English.

XX Synthetic peptides based on portions of cedar pollen allergens A

XX (AAR81586) and B (AAR81587) were tested for their ability to activate

XX cedar allergen-specific T-cells, but not allergen-specific IgE

XX antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell

XX epitopes. These peptides, plus subsequences (AAR81573-79) essential

XX for T-cell recognition, and homologous peptides (AAR81588-96) can

XX be used as immunotherapeutic agents to treat or prevent cedar

XX pollinosis, avoiding side-effects such as anaphylaxis.

SO Sequence 353 AA;

Query Match 100.0%; Score 83; DB 17; Length 353;

Best Local Similarity 100.0%; Pred. No. 1.2e-06; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMPMYIAGYKTFDGR 15

DB 81 KMPMYIAGYKTFDGR 95

RESULT 8

AAV25664 standard; protein; 373 AA.

XX AAV25664;

XX 30-SEP-1999 (first entry)

XX Cedar allergen 493634 Cry j IB protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;

XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;

XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;

XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;

XX cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;

XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

XX Cedrus sp.

XX WO934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 73; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a

XX polypeptide allergen and comprises administering to the patient a peptide

XX derived from the allergen where restriction to a MHC Class II molecule

XX possessed by the patient can be demonstrated for the peptide and the

XX peptide is able to induce a late phase response in an individual who

XX possesses the MHC Class II molecule. The methods can be used for

XX desensitizing patients to allergens present in e.g. grass, tree and weed

XX (including ragweed) pollens, fungi and moulds, foods, stinging insects,

XX the chironomidae (non-biting midges), spiders and mites, housefly, fruit

XX fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,

XX non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of

XX Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,

XX sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to

XX produce immunological vaccines which may be used to prevent and/or treat

XX conditions involving hypersensitivity to allergens. This sequence

SO Sequence 373 AA;

Query Match 100.0%; Score 83; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 1.3e-06; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMPMYIAGYKTFDGR 15

DB 101 KMPMYIAGYKTFDGR 115

RESULT 9

AAV25668 standard; protein; 373 AA.

XX AAV25668;

```

XX 30-SEP-1999 (first entry)
DT Japanese cedar allergen 541803 Cry j I precursor protein fragment.
DE
XX
XX Major histocompatibility complex; class II; desensitizing; human;
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; string;
XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
XX cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.
XX
XX Cedrus sp.
OS
XX WO9934926-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GB00080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
XX
XX MPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 75; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
XX polypeptide allergen and comprises administering to the patient a peptide
XX derived from the allergen where restriction to a MHC Class II molecule
XX possessed by the patient can be demonstrated for the peptide and the
XX peptide is able to induce a late phase response in an individual who
XX possesses the MHC Class II molecule. The methods can be used for
XX desensitizing patients to allergens present in e.g. grass, tree and weed
XX (including ragweed) pollens, fungi and moulds, foods, stinging insects,
XX the chironomidae (non-biting midges), spiders and mites, housefly, fruit
XX fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
XX non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
XX tendrillo mottor beetle, mammals such as cat, dog, horse, cow, pig,
XX sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
XX produce immunological vaccines, which may be used to prevent and/or treat
XX conditions involving hypersensitivity to allergens. This sequence
XX represents a Japanese cedar (Cedrus sp.) allergen 541803 Cry j I
XX precursor.
XX
XX Sequence 373 AA;
XX
XX Query Match 100.0%; Score 83; DB 20; Length 373;
XX Best Local Similarity 100.0%; Pred. No. 1.3e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 KMPWYIAGYKTFDGR 15
XX |||||
XX DB 101 KMPWYIAGYKTFDGR 115
XX
XX RESULT 10
XX ID AAR31937 standard; Protein; 374 AA.
XX
XX AAR31937;
XX
XX 03-JUN-1993 (first entry)
XX
XX Cry j I.
XX
XX Japanese cedar pollen; allergen; antigen; allergy; B cell; T cell.

```

```

XX OS Cryptomeria japonica.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT Protein /note= "signal peptide"
XX FT /note= "mature Cry j I"
XX
XX PN WO9301213-A.
XX
XX 21-JAN-1993.
XX
XX 10-JUL-1992; 92WO-US05661.
XX
XX 12-JUL-1991; 91US-0729134.
XX
XX 15-JUL-1991; 91US-0730452.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Bond JF, Griffith J, Pollock J;
XX
XX MPI; 1993-045434/05.
XX
XX N-PSDB; AAQ35304.
XX
XX Nucleic acid sequence encoding Cryptomeria japonica allergen -
XX for the diagnosis treatment and prevention of allergic reactions
XX to Japanese cedar pollen
XX
XX Claim 11; Page 42; 69pp; English.
XX
XX CC Fresh pollen and staminate cone samples were collected from a single
XX Cryptomeria japonica (Japanese cedar) tree. RNA was prep'd. and used
XX to synthesize cDNA. The cDNA was subjected to successive rounds of
XX PCR to yield a full length Cry j I clone. Cry j I or an antigenic
XX fragment of it may be used for detecting, treating and preventing an
XX allergic response to Japanese cedar pollen allergen. It is capable of
XX modifying both the B and T cell response to Cry j I and T cell response
XX to a Cry j I antigen.
XX
XX SQ Sequence 374 AA;
XX
XX Query Match 100.0%; Score 83; DB 14; Length 374;
XX Best Local Similarity 100.0%; Pred. No. 1.3e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 KMPWYIAGYKTFDGR 15
XX |||||
XX DB 102 KMPWYIAGYKTFDGR 116
XX
XX RESULT 11
XX ID AAR45541 standard; Protein; 374 AA.
XX
XX AAR45541;
XX
XX 13-JUL-1994 (first entry)
XX
XX Cry j I pollen allergen.
XX
XX Japanese cedar; detection; allergy; treatment; diagnosis;
XX T cell epitope; sensitivity.
XX
XX OS Cryptomeria japonica.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /note= "signal peptide"
XX FT /note= "mature peptide"
XX
XX PN WO9401560-A.

```

```

XX 20-JAN-1994.
PD
XX
PF 15-JAN-1993; 93WO-US00139.
PR 01-SEP-1992; 92US-0938990.
PR 10-JUL-1992; 92WO-US05661.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J;
XX WPI; 1994-035066/04.
DR N-PSDB; AAQ55271.
XX
PT Antigens derived from Japanese cedar pollen allergen Cry j I -
PT contain at least two T cell epitope(s), used to treat or diagnose
PT allergy
XX
PS Disclosure; Fig 4; 137pp; English.
XX
CC The sequence is that of the Japanese cedar pollen allergen
CC Cry j I which contains at least two T cell epitopes. Peptide
CC antigens derived from it can be used for the treatment and
CC diagnosis of allergies associated with Japanese cedar pollen.
CC The peptides have enhanced therapeutic properties but reduced
CC side effects compared to naturally occurring allergens.
XX
SQ Sequence 374 AA;
XX
Query Match 100.0%; Score 83; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KMPWYIAGYKTFDGR 15
DB 102 KMPWYIAGYKTFDGR 116
XX
RESULT 12
AAR60166
ID AAR60166 standard; Protein; 374 AA.
XX
AC AAR60166;
XX
DT 24-MAR-1995 (first entry)
XX
DE Japanese cedar pollen antigen CryjI.
XX
KW Japanese cedar; pollen antigen; allergen; CryjI; sugi; pollinosis.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= signal_peptide
FT Protein 22..374
FT /label= mature_CryjI
XX
XX JPO6197768-A.
XX
XX 19-JUL-1994.
XX
XX 07-JAN-1993; 93JP-0001116.
XX
XX 07-JAN-1993; 93JP-0001116.
XX
XX (MEIJ) MEIJI SEIKA KAISHA.
XX
XX WPI; 1994-268680/33.
XX
XX N-PSDB; AAQ71601.
XX
PT Sugi (Japanese cedar) pollen antigen CryjI - is useful for

```

```

PT diagnosis, treatment and prevention of sugi pollinosis
XX
XX Claim 2; Page 5-7; 9pp; Japanese.
XX
XX The coding sequence for the Japanese cedar ("sugi") pollen allergen
XX Cry j I was isolated from a cDNA library prepared from polyA mRNA. All
XX or part of the CryjI protein can be used for diagnosis, treatment
XX and prevention of sugi pollinosis.
XX
SQ Sequence 374 AA;
XX
Query Match 100.0%; Score 83; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KMPWYIAGYKTFDGR 15
DB 102 KMPWYIAGYKTFDGR 116
XX
RESULT 13
AAR82490
ID AAR82490 standard; Protein; 374 AA.
XX
AC AAR82490;
XX
DT 15-APR-1996 (first entry)
XX
DE Cry j I Japanese Cedar pollen allergen.
XX
KW Cry j I; Japanese cedar pollen allergen; modified; drug production;
KW allergy; Cryptomeria japonica.
XX
OS Cryptomeria japonica.
XX
PN WO9527786-A1.
XX
PD 19-OCT-1995.
XX
PF 06-APR-1995; 95WO-US04249.
XX
PR 06-DEC-1994; 94US-0350225.
PR 08-APR-1994; 94US-0226248.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;
PI Shaked Z;
XX
DR WPI; 1995-366391/47.
DR N-PSDB; AAT04248.
XX
PT Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for
PT treating allergy to Japanese cedar pollen allergen or
PT immunologically cross reactive allergens
XX
PS Disclosure; Figure 1; 60pp; English.
XX
XX Novel peptides of cry j I have been modified as a part of a
XX preformulation scheme to develop an optimised drug product for
XX therapeutic treatment of humans suffering from allergy to Japanese
XX cedar pollen allergen or an allergen which is immunologically cross
XX reactive with Japanese cedar pollen allergen. Such modified peptides
XX possess certain characteristics which render them particularly
XX suitable for drug product formulation. Peptide fragments of Cry j I,
XX modified and unmodified, are given in AAR82491-R82525.
XX
SQ Sequence 374 AA;
XX
Query Match 100.0%; Score 83; DB 16; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



OY 1 KMPMYIAGYKTFDGR 15  
 DB 102 KMPMYIAGYKTFDGR 116

## RESULT 14

AAV25665 standard; protein; 374 AA.

AAV25665;

30-SEP-1999 (first entry)

Cedar allergen 493632 Cry j IA protein fragment.

Major histocompatibility complex; class II; desensitizing; human;  
 allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

Cedrus sp.

WO9934826-A1.

15-JUL-1999.

11-JAN-1999; 99WO-GB00080.

21-SEP-1998; 98GB-0020474.

09-JAN-1998; 98GB-0000445.

(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

Kay AB, Larche M;

WPI; 1999-458255/38.

Desensitizing patients to polypeptide allergens

Example 6; Page 73; 117pp; English.

This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of tenbrito mollitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a cedar (Cedrus sp.) allergen 493632 Cry j IA.

Sequence 374 AA;

Query Match 100.0%; Score 83; DB 20; Length 374;

Best Local Similarity 100.0%; Pred. No. 1.3e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 15

## RESULT 15

AAV25669

ID AAV25669 standard; protein; 374 AA.

AAV25669;

30-SEP-1999 (first entry)

Japanese cedar allergen 541802 Cry j I precursor protein fragment.

Major histocompatibility complex; class II; desensitizing; human;  
 allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

Cedrus sp.

WO9934826-A1.

15-JUL-1999.

11-JAN-1999; 99WO-GB00080.

21-SEP-1998; 98GB-0020474.

09-JAN-1998; 98GB-0000445.

(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

Kay AB, Larche M;

WPI; 1999-458255/38.

Desensitizing patients to polypeptide allergens

Example 6; Page 75; 117pp; English.

This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of tenbrito mollitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a Japanese cedar (Cedrus sp.) allergen 541802 Cry j I precursor.

Sequence 374 AA;

Query Match 100.0%; Score 83; DB 20; Length 374;

Best Local Similarity 100.0%; Pred. No. 1.3e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KMPMYIAGYKTFDGR 15

DB 102 KMPMYIAGYKTFDGR 116

Search completed: April 20, 2003, 13:06:07  
 Job time: 19.1974 secs





CC AAR97871-R97960 are overlapping peptides used for the epitope mapping  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.

XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 17; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDGIIAAYONPASWK 15  
 DB 1 VDGIIAAYONPASWK 15

# RESULT 2

ID AAM57760 standard; peptide; 15 AA.

AC AAM57760;

DT 17-SEP-1998 (first entry)

DE Residues 66-80 of Cry j 2.

XX Cry j 2; Japanese cedar pollen antigen; allergy; immunotherapy;

KM HLA class II molecule.

XX Cryptomeria japonica.

PN WC9820902-A1.

PD 22-MAY-1998.

PF 12-NOV-1997; 97WO-JP04129.

PR 13-NOV-1996; 96JP-0302053.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kuno K, Kume A, Some T;

DR WPI; 1998-297617/26.

XX Peptides derived from Japanese cedar pollen antigens are

PT immunotherapeutic agents - useful for allergy treatment and typing

PT HLA class II molecules in allergy sufferers

PS Claim 12; Page 30; 50pp; Japanese.

XX This sequence represents residues 66-80 of the Cry j 2 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.

XX Sequence 15 AA;

Query Match 100.0%; Score 81; DB 19; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDGIIAAYONPASWK 15  
 DB 1 VDGIIAAYONPASWK 15

# RESULT 3

ID AAR81580 standard; peptide; 17 AA.

AC AAR81580;

DT 24-MAY-1996 (first entry)

DE Cedar pollen allergen peptide 8 (T-cell epitope).

XX Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;

KW IGE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.

XX Synthetic.

PN EP700929-A2.

PD 13-MAR-1996.

PF 08-SEP-1995; 95EP-0306295.

PR 14-JUL-1995; 95JP-0200221.

PR 10-SEP-1994; 94JP-0242137.

PR 14-JUL-1995; 95JP-0200204.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Hino K, Saito S, Taniguchi Y;

DR WPI; 1996-140976/15.

XX New peptide(s) derived from cedar pollen allergens - activate

PT allergen-specific T-cells, but not allergen-specific IGE antibodies,

PT used for treating cedar pollinosis

PS Claim 4; Page 28; 36pp; English.

XX Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IGE  
 CC antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81573-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.

XX Sequence 17 AA;

Query Match 100.0%; Score 81; DB 17; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDGIIAAYONPASWK 15  
 DB 2 VDGIIAAYONPASWK 16

# RESULT 4

ID AAW80345 standard; peptide; 17 AA.

AC AAW80345;

DT 11-JAN-1999 (first entry)

DE Sugi allergen protein Cryj2 derived epitope for T cells.

XX T cell epitope; sugi allergen proteins Cryj1, Cryj2; treatment;

KW sugi-pollinosis; allergic reaction; pollen.  
 CC Synthetic.  
 OS  
 XX JP10259198-A.  
 PN  
 XX 29-SEP-1998.  
 PD  
 XX 22-DEC-1997; 97JP-0353448.  
 PF  
 XX 24-DEC-1996; 96JP-0343441.  
 PR  
 XX  
 XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY ) SANKYO CO LTD.  
 DR WPI; 1998-577037/49.  
 XX  
 XX A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 XX  
 XX Claim 7; Page 18; 21pp; Japanese.  
 PS  
 CC AAM80339-58 represent epitopes for T cells, derived from the sugi  
 CC allergen proteins Cryj1 (AAM80339-44, AAM80350-53 and AAM80356-58) and  
 CC Cryj2 (AAM80345-49 and AAM80354-55). The peptides are useful for the  
 CC treatment of sugi-pollinosis, an allergic reaction of the body to  
 CC pollen.  
 CC  
 XX Sequence 17 AA;  
 SQ  
 Query Match 100.0%; Score 81; DB 19; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VDGIIAAYQNPAWSK 15  
 2 VDGIIAAYQNPAWSK 16  
 DB  
 RESULT 5  
 AAM27369  
 ID AAM27369 standard; peptide; 80 AA.  
 XX  
 AC AAM27369;  
 XX  
 DT 24-MAR-1998 (first entry)  
 XX  
 DE Multi-epitope peptide used as immunotherapeutic agent #1.  
 XX  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 OS  
 XX WO9732600-A1.  
 PN  
 XX 12-SEP-1997.  
 PD  
 XX 10-MAR-1997; 97WO-JP00740.  
 PF  
 XX 10-MAR-1996; 96JP-0080702.  
 PR  
 XX (MEIJ ) MEIJI MILK PROD CO LTD.  
 PA  
 XX Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 PI WPI; 1997-470495/43.  
 DR  
 XX Peptide immuno:therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 XX  
 PS Claim 6; Page 31; 58pp; Japanese.

XX The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 XX  
 XX Sequence 80 AA;  
 SQ  
 Query Match 100.0%; Score 81; DB 18; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VDGIIAAYQNPAWSK 15  
 66 VDGIIAAYQNPAWSK 80  
 DB  
 RESULT 6  
 AAM27370  
 ID AAM27370 standard; peptide; 105 AA.  
 XX  
 AC AAM27370;  
 XX  
 DT 24-MAR-1998 (first entry)  
 XX  
 DE Multi-epitope peptide used as immunotherapeutic agent #2.  
 XX  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 OS  
 XX WO9732600-A1.  
 PN  
 XX 12-SEP-1997.  
 PD  
 XX 10-MAR-1997; 97WO-JP00740.  
 PF  
 XX 10-MAR-1996; 96JP-0080702.  
 PR  
 XX (MEIJ ) MEIJI MILK PROD CO LTD.  
 PA  
 XX Dairiki K, Iwama A, Kino K, Kume A, Sone T;  
 PI WPI; 1997-470495/43.  
 DR  
 XX Peptide immuno:therapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 XX  
 PS Claim 6; Page 31; 58pp; Japanese.  
 XX  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 XX  
 XX Sequence 105 AA;  
 SQ  
 Query Match 100.0%; Score 81; DB 18; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAYONPASWK 15  
 DB 91 VDGIIAAYONPASWK 105

RESULT 7  
 AAM27371

ID AAM27371 standard; peptide; 134 AA.

AC AAM27371;

DT 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #3.

KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

PN WO9732600-A1.

PD 12-SEP-1997.

PF 10-MAR-1997; 97WO-JP00740.

PR 10-MAR-1996; 96JP-0080702.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;

DR WPI; 1997-470495/43.

PT Peptide immuno-therapeutic agent to treat allergic diseases -

PT contains multi-epitope peptide containing T cell epitope regions

PT from different allergens

PS Claim 6; Page 32; 58pp; Japanese.

CC The present sequence represents a multi-epitope peptide which is used as

CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2

CC or more different allergens (preferably linked via arginine or lysine

CC dimers), where the T cell epitope regions have a positivity index

CC greater than 100 as measured in a patient group responding to the

CC allergen; have at least 70% reactivity with lymphocytes from patients

CC responding to the allergen; and are not reactive with immunoglobulin E

CC (IgE) antibodies from patients responsive to the allergen. The agent can

CC be used to prevent and treat a wide variety of allergic diseases, e.g. by

CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

CC

CC

CC

CC

CC

CC

CC

XX Japonicum allergen; residues 55-514; induced histamine release;

KW antiallergic peptide; IgE cross-linking inhibition.

OS Japonicum sp.

PN WO9502412-A.

PD 26-JAN-1995.

PF 15-JUL-1994; 94WO-JP01164.

PR 16-JUL-1993; 93JP-0177008.

PR 01-SEP-1993; 93JP-0217725.

PR 07-APR-1994; 94JP-0069336.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Kino K, Kohno Y, Komiyama N, Sone T;

DR WPI; 1995-067159/09.

DR N-PSDB; AAQ84044.

PT Peptide antiallergic agent - inhibits cross-linking of allergen

PT with IgE antibody

PS Disclosure; Pages 26-27; 46pp; Japanese.

CC AAQ84044 encodes AAR69791 Japonicum allergen residues 55-514, from

CC which the antiallergic peptides AAR69845-R69809 were derived.

CC The peptides ability to inhibit the cross-linking of an allergen,

CC to an IgE antibody can be used in the prevention and treatment of

CC allergic diseases.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

```

XX Allergic Cry j II protein and fragments from Japanese cedar
PT pollen - used to diagnose, treat and prevent Japanese cedar
PT pollinosis
XX
PS Claim 2; Fig 4; 89pp; English.
XX
CC The sequence is of a Japanese cedar pollen allergen Cry j
CC II. The protein and its fragments can be used for diagnosis and
CC treatment of Japanese cedar pollinosis and to identify similar
CC sequences in other plants.
CC See also AAR53692-6.
XX
SQ Sequence 514 AA;

Query Match 100.0%; Score 81; DB 15; Length 514;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAAYQNPASWK 15
DB 120 VDGIIAAAYQNPASWK 134

RESULT 10
AAR74333
ID AAR74333 standard; Protein; 514 AA.
XX
AC AAR74333;
XX
DT 01-NOV-1995 (first entry)
XX
DE Japanese cedar pollen allergen.
XX
KM Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;
KM desensitizer.
XX
OS Cryptomeria japonica.
XX
PN EP655500-A.
XX
PD 31-MAY-1995.
XX
PF 03-NOV-1994; 94EP-0308117.
XX
PR 27-DEC-1993; 93JP-0346814.
PR 05-NOV-1993; 93JP-0299151.
PR 20-DEC-1993; 93JP-0344596.
XX
PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.
XX
PI Kurimoto M, Namba M, Torigoe K;
XX
DR WPI; 1995-195588/26.
DR N-PSDB; AAO90156.
XX
PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it,
PT useful for treatment and diagnosis of cedar pollen allergy
XX
PS Claim 5; Page 26-28; 41pp; English.
XX
CC The gene encoding an allergen of Japanese cedar pollen was isolated
CC by PCR amplification using primers based on portions of the allergen
CC protein. The gene was used for recombinant allergen production in
CC E. coli (vector plasmid pKK-223-3).
XX
SQ Sequence 514 AA;

Query Match 100.0%; Score 81; DB 16; Length 514;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAAYQNPASWK 15
DB 120 VDGIIAAAYQNPASWK 134

```

```

DB 120 VDGIIAAAYQNPASWK 134

RESULT 11
AAR69792
ID AAR69792 standard; Protein; 514 AA.
XX
AC AAR69792;
XX
DT 27-SEP-1995 (first entry)
XX
DE Japonicum allergen.
XX
KM Japonicum allergen; induced histamine release; antiallergic peptide;
KM IGE cross-linking inhibition.
XX
OS Japonicum sp.
XX
PN WO9502412-A.
XX
PD 26-JAN-1995.
XX
PF 15-JUL-1994; 94WO-JP01164.
XX
PR 16-JUL-1993; 93JP-0177008.
PR 01-SEP-1993; 93JP-0217725.
PR 07-APR-1994; 94JP-0069336.
XX
PA (MEIP ) MEIJI MILK PROD CO LTD.
XX
PI Kino K, Kohno Y, Komiyama N, Sone T;
XX
DR WPI; 1995-067159/09.
DR N-PSDB; AAO84045; AAO84046.
XX
PT Peptide antiallergic agent - inhibits cross-linking of allergen
PT with IGE antibody
XX
PS Example 3; Pages 27-28; 46pp; Japanese.
XX
CC AAO84045 encodes AAR69792 Japonicum allergen, from which the
CC antiallergic peptides AAR69845-R69809 were derived. The peptides ability
CC to inhibit the cross-linking of an allergen, to an IGE antibody can be
CC used in the prevention and treatment of allergic diseases.
XX
SQ Sequence 514 AA;

Query Match 100.0%; Score 81; DB 16; Length 514;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAAYQNPASWK 15
DB 120 VDGIIAAAYQNPASWK 134

RESULT 12
AAR93599
ID AAR93599 standard; Protein; 514 AA.
XX
AC AAR93599;
XX
DT 16-AUG-1996 (first entry)
XX
DE Japan cedar pollen Cry j II allergen.
XX
KM Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;
KM Sugi pollinosis; diagnosis; treatment.
XX
OS Cryptomeria japonica.
XX
FH Key Location/Qualifiers

```

```

FT Peptide 1.54
FT /Label= sig_peptide
FT Protein 55.514
FT /*label= mac_protein
XX
XX JP08047392-A.
XX
XX 20-FEB-1996.
XX
XX 07-NOV-1994; 94JP-0297840.
XX
XX 26-MAY-1994; 94JP-0134868.
XX
XX 05-NOV-1993; 93JP-0276773.
XX
XX (MEIP ) MEIJI MILK PROD CO LTD.
XX
XX WPI; 1996-166249/17.
XX
XX N-PSDB; AAT18102.
XX
XX Part of specified 460 amino acid protein
XX
XX Claim 1; Page 10-11; 17pp; Japanese.
XX
XX AAR3599 is a Japan cedar pollen Cry j II allergen which is useful
XX in the diagnosis, prevention and treatment of Sugi pollinosis,
XX the allergic reaction to Japan cedar pollen. Significant regions of
XX the allergen were identified using overlapping peptides of the full
XX epitope derived from a Cry j II antigen-specific T cell line
XX (see AAR97871-R97960). Amino acids 66-80 (AAR97884) and 186-200
XX (R978908) of the full mature 460 amino acid allergen are the most
XX allergenic of the 90 peptides tested.
XX
XX Sequence 514 AA;
XX
XX Query Match 100.0%; Score 81; DB 17; Length 514;
XX Best Local Similarity 100.0%; Pred. No. 1e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VDGIIAAYQNPASWK 15
XX
XX 120 VDGIIAAYQNPASWK 134
XX
XX
XX RESULT 13
XX AAR81586
XX ID AAR81586 standard; Protein; 514 AA.
XX
XX AAR81586;
XX
XX 24-MAY-1996 (first entry)
XX
XX Cedar pollen allergen A.
XX
XX Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;
XX antibody; pollinosis; therapy; immunotherapy.
XX
XX Cryptomeria japonica.
XX
XX EP700929-A2.
XX
XX 13-MAR-1996.
XX
XX 08-SEP-1995; 95EP-0306295.
XX
XX 14-JUL-1995; 95JP-0200221.
XX
XX 10-SEP-1994; 94JP-0242137.
XX
XX 14-JUL-1995; 95JP-0200204.
XX
XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.
XX
XX Hino K, Saito S, Taniguchi Y;
XX

```

```

DR WPI; 1996-140976/15.
XX
XX New peptide(s) derived from cedar pollen allergens - activate
XX allergen-specific T-cells, but not allergen-specific IgE antibodies,
XX used for treating cedar pollinosis
XX
XX Claim 5; Page 29-30; 36pp; English.
XX
XX
XX Synthetic peptides based on portions of cedar pollen allergens A
XX (AAR81586) and B (AAR81587) were tested for their ability to activate
XX cedar allergen-specific T-cells, but not allergen-specific IgE
XX antibodies. 6 peptides (AAR81580-R81585) were identified as T-cell
XX epitopes. These peptides, plus subsequences (AAR81573-79) essential
XX for T-cell recognition, and homologous peptides (AAR81588-96) can
XX be used as immunotherapeutic agents to treat or prevent cedar
XX pollinosis, avoiding side-effects such as anaphylaxis.
XX
XX Sequence 514 AA;
XX
XX
XX Query Match 100.0%; Score 81; DB 17; Length 514;
XX Best Local Similarity 100.0%; Pred. No. 1e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VDGIIAAYQNPASWK 15
XX
XX 120 VDGIIAAYQNPASWK 134
XX
XX
XX RESULT 14
XX AAY25666
XX ID AAY25666 standard; protein; 514 AA.
XX
XX AAY25666;
XX
XX 30-SEP-1999 (first entry)
XX
XX Japanese cedar allergen 1076242 Cry j II precursor protein fragment.
XX
XX Major histocompatibility complex; class II; desensitizing; human;
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; string;
XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
XX cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.
XX
XX Cedrus sp.
XX
XX WO9934826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GB00080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 74; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
XX polypeptide allergen and comprises administering to the patient a peptide
XX derived from the allergen where restriction to a MHC class II molecule
XX possessed by the patient can be demonstrated for the peptide and the
XX peptide is able to induce a late phase response in an individual who
XX possesses the MHC class II molecule. The methods can be used for
XX desensitizing patients to allergens present in e.g. grass, tree and weed
XX

```



CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrito molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II  
 CC precursor.

XX  
 XX Sequence 514 AA;  
 SQ

Query Match 100.0%; Score 81; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAAYQNPASWK 15  
 |||||  
 DB 120 VDGIIAAAYQNPASWK 134

## RESULT 15

AAV25667  
 ID AAV25667 standard; protein; 514 AA.

XX AAV25667;  
 AC

DT 30-SEP-1999 (first entry)

DE Japanese cedar allergen 1076241 Cry j II protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

XX Cedrus sp.  
 OS

XX WO9934826-A1.  
 PN

XX 15-JUL-1999.  
 PD

XX 11-JAN-1999; 99WO-GB00080.  
 PF

XX 21-SEP-1998; 98GB-0020474.  
 PR

XX 09-JAN-1998; 98GB-0000445.  
 PR

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 PA

XX Kay AB, Larche M;  
 PI

XX WPI; 1999-458255/38.  
 DR

XX Desensitizing patients to polypeptide allergens  
 PT

XX Example 6; Page 74; 117p; English.  
 PS

CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitising patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC Tenbrito molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to

CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a Japanese cedar (Cedrus sp.) allergen 1076241 Cry j II.

XX  
 XX Sequence 514 AA;  
 SQ

Query Match 100.0%; Score 81; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDGIIAAAYQNPASWK 15  
 |||||  
 DB 120 VDGIIAAAYQNPASWK 134

Search completed: April 20, 2003, 13:06:09  
 Job time : 19.1974 Secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05 ; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524D-36  
Perfect score: 80  
Sequence: 1 PCVFKRVSNVITHG 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	80	100.0	15	AA89291	Japanese cedar pol
2	80	100.0	20	AA84552	Cry j I pollen all
3	80	100.0	20	AA82501	Cry j I Japanese C
4	80	100.0	30	AA44683	T-cell epitope pep
5	80	100.0	353	AA87538	Japanese cedar pol
6	80	100.0	353	AA81587	Cedar pollen aller
7	80	100.0	373	AA25664	Cedar allergen 493
8	80	100.0	374	AA25668	Japanese cedar all
9	80	100.0	374	AA81937	Cry j I. Cryptome
10	80	100.0	374	AA845541	Cry j I pollen all

11	80	100.0	374	AA86016	Japanese cedar pol
12	80	100.0	374	AA882490	Cry j I Japanese C
13	80	100.0	374	AA25665	Japanese cedar 493
14	80	100.0	374	AA25669	Japanese cedar all
15	73	91.2	15	AA57750	Residues 106-120 o
16	64	80.0	13	AA884112	Amino acid sequenc
17	64	80.0	36	AA884102	Immunomodulatory p
18	64	80.0	80	AA827369	Multi-epitope pep
19	64	80.0	105	AA827370	Multi-epitope pep
20	64	80.0	134	AA827371	Multi-epitope pep
21	61	76.2	370	AA84578	Multi-epitope pep
22	58	72.5	20	AA84579	Multi-epitope pep
23	58	72.5	354	AA842133	T-cell epitope pep
24	58	72.5	354	AA842133	Chamaecyparis obtu
25	58	72.5	354	AA842121	Japanese cypress p
26	57	71.2	264	AA839243	Chamaecyparis obtu
27	57	71.2	270	AA839243	Arabiopsis thalia
28	57	71.2	274	AA839243	Arabiopsis thalia
29	57	71.2	280	AA839242	Arabiopsis thalia
30	57	71.2	384	AA839241	Arabiopsis thalia
31	57	71.2	384	AB890972	Arabiopsis thalia
32	57	71.2	390	AA829241	Herbicidally activ
33	56	70.0	36	AA827996	Arabiopsis thalia
34	56	70.0	45	AA827995	Arabiopsis thalia
35	56	70.0	266	AA839368	Arabiopsis thalia
36	56	70.0	272	AA842704	Arabiopsis thalia
37	56	70.0	274	AA832458	Arabiopsis thalia
38	56	70.0	276	AA839367	Arabiopsis thalia
39	56	70.0	349	AA832457	Arabiopsis thalia
40	56	70.0	394	AA832456	Arabiopsis thalia
41	56	70.0	394	AB893044	Arabiopsis thalia
42	56	70.0	394	AB893045	Herbicidally activ
43	56	70.0	404	AA842703	Herbicidally activ
44	56	70.0	404	AB893086	Herbicidally activ
45	56	70.0	406	AA806909	Arabiopsis thalia

## ALIGNMENTS

RESULT 1

AA89291 ID AA89291 standard; peptide; 15 AA.

AA89291; AC AA89291; XX 12-MAR-1996 (first entry)

XX Japanese cedar pollen allergen Cryj I derived T-cell epitope peptide.

DE Japanese cedar; pollen allergen; Cryj I; T-cell epitope; peptides;

XX prevention; treatment; cryptomeria pollinosis.

KW Cryptomeria japonica.

XX JP07118295-A.

PN 09-MAY-1995.

XX 20-OCT-1993; 93JP-0262626.

XX 20-OCT-1993; 93JP-0262626.

XX 20-OCT-1993; 93JP-0262626.

XX (MEIP) MEIJI MILK PROD CO LTD.

PA WPI; 1995-203834/27.

DR New cryptomeria pollen allergen T-cell epitope peptide - used for

XX prevention, treatment and investigation of Japanese cedar pollinosis

PT Claim 5; Page 2; 8pp; Japanese.

PS AA875388 is the Japanese cedar pollen allergen Cryj I, from which the

CC

CC T-cell epitope peptides AAR82989-R82995 were derived. The peptides  
CC can be used for the prevention and treatment of Cryptomeria  
CC pollenosis, and also for the investigation of pollenosis.

SQ Sequence 15 AA;

Query Match 100.0%; Score 80; DB 16; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVIHG 15  
DB 1 PCVFIRKRVSNVIHG 15

RESULT 2  
AAR45552  
ID AAR45552 standard; Protein; 20 AA.

AC AAR45552;

DT 13-JUN-1994 (first entry)

DE Cry j I pollen allergen peptide CJI-11.

KM Japanese cedar; detection; allergy; treatment; diagnosis;

KW T cell epitope; sensitivity.

OS Cryptomeria japonica.

XX WO9401560-A.

XX 20-JAN-1994.

XX 15-JAN-1993; 93WO-US00139.

XX 01-SEP-1992; 92US-0938990.

XX 10-JUL-1992; 92WO-US05661.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Bond UF, Garman RD, Griffith IJ, Kuo M, Pollock J;

XX WPI, 1994-035066/04.

PT Antigens derived from Japanese cedar pollen allergen Cry j I -

PT contain at least two T cell epitope(s), used to treat or diagnose

PT allergy

XX Claim 1; Fig 13; 137pp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar

CC pollen allergen Cry j I (amino acids 101-120). The peptide, CJI-11,

CC can be used for the treatment and diagnosis of allergies associated

CC with Japanese cedar pollen. It has enhanced therapeutic properties

CC but reduced side effects compared to naturally occurring allergens.

XX Sequence 20 AA;

Query Match 100.0%; Score 80; DB 15; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVIHG 15  
DB 6 PCVFIRKRVSNVIHG 20

RESULT 3  
AAR82501  
ID AAR82501 standard; Protein; 20 AA.  
AC AAR82501;

XX 15-APR-1996 (first entry)

DT Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-11).

DE Cry j I; Japanese cedar pollen allergen; modified; drug production;

XX allergy; Cryptomeria japonica.

OS Cryptomeria japonica.

XX WO9527786-A1.

XX 19-OCT-1995.

XX 06-APR-1995; 95WO-US04249.

XX 06-DEC-1994; 94US-0350225.

XX 08-APR-1994; 94US-0226248.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Chen X, Evans S, Franzen HM, Kuo M, Powers SP;

XX Shaked Z;

XX WPI, 1995-366391/47.

XX Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for

XX treating allergy to Japanese cedar pollen allergen or

XX immunologically cross reactive allergens

XX Disclosure; Figure 2; 60pp; English.

XX Novel peptides of cry j I have been modified as a part of a

XX preformulation scheme to develop an optimised drug product for

XX therapeutic treatment of humans suffering from allergy to Japanese

XX cedar pollen allergen or an allergen which is immunologically cross

XX reactive with Japanese cedar pollen allergen. Such modified peptides

XX possess certain characteristics which render them particularly

XX suitable for drug product formulation. Peptide fragments of Cry j I,

XX modified and unmodified, are given in AAR82491-R82525. This peptide

XX fragment corresponds to amino acids 101-120 of the allergen mature

XX protein.

XX Sequence 20 AA;

Query Match 100.0%; Score 80; DB 16; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PCVFIRKRVSNVIHG 15  
DB 6 PCVFIRKRVSNVIHG 20

RESULT 4  
AAW44683  
ID AAW44683 standard; peptide; 30 AA.  
AC AAW44683;  
XX 01-MAY-1998 (first entry)  
DT T-cell epitope peptide #2 of sugi pollen antigen.  
DE T-cell epitope; sugi pollen antigen; sugi pollinosis.  
XX Synthetic.  
XX Cryptomeria japonica.  
XX JPI0007700-A.  
XX 13-JAN-1998.

PF 24-JUN-1996; 96JP-0163287.  
 XX 24-JUN-1996; 96JP-0163287.  
 XX (DAIL ) DAICEL CHEM IND LTD.  
 PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
 XX WPI, 1998-133630/13.  
 DR T cell epitope peptide of sugi pollen antigen - useful in the  
 PT treatment of sugi pollinosis  
 XX  
 PS Claim 1; Page 4; 14pp; Japanese.  
 XX  
 CC T-cell epitope peptides AAW4682-88 and their derivatives react with  
 CC sugi pollinosis patient peripheral blood T lymphocytes. A composition  
 CC prepared by combining at least 2 of the above peptides and/or their  
 CC derivatives is used for the prevention and treatment of sugi  
 CC pollinosis.  
 CC  
 SQ Sequence 30 AA;  
 Query Match 100.0%; Score 80; DB 19; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PCVFIRKRVSNVTHG 15  
 DB 6 PCVFIRKRVSNVTHG 20  
 RESULT 5  
 AAR75388 standard; protein; 353 AA.  
 AC AAR75388;  
 XX  
 DT 12-MAR-1996 (first entry)  
 XX  
 DE Japanese cedar pollen allergen Cry I.  
 XX  
 KM Japanese cedar pollen allergen; Cry I; T-cell epitope; peptides;  
 KM prevention; treatment; cryptomeria pollinosis.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 61..75  
 FT Peptide /note="T-cell epitope peptide"  
 FT Peptide 91..105  
 FT Peptide /note="T-cell epitope peptide"  
 FT Peptide 106..120  
 FT Peptide /note="T-cell epitope peptide"  
 FT Peptide 146..160  
 FT Peptide /note="T-cell epitope peptide"  
 FT Peptide 211..225  
 FT Peptide /note="T-cell epitope peptide"  
 FT Peptide 326..340  
 FT Peptide /note="T-cell epitope peptide"  
 FT Peptide 335..346  
 FT Peptide /note="T-cell epitope peptide"  
 XX  
 PM JP07118295-A.  
 XX  
 PD 09-MAY-1995.  
 XX  
 PF 20-OCT-1993; 93JP-0262626.  
 XX  
 PR 20-OCT-1993; 93JP-0262626.  
 XX  
 PA (MEIJ ) MEIJI MILK PROD CO LTD.  
 XX  
 DR WPI, 1995-203634/27.

XX  
 PT New cryptomeria pollen allergen T-cell epitope peptide - used for  
 PT prevention, treatment and investigation of Japanese cedar pollinosis  
 XX  
 PS Disclosure; Figs 1-2; 8pp; Japanese.  
 XX  
 CC AAR75388 is the Japanese cedar pollen allergen Cry I, from which the  
 CC T-cell epitope peptides AAR8289-R8295 were derived. The peptides  
 CC can be used for the prevention and treatment of cryptomeria  
 CC pollinosis, and also for the investigation of pollinosis.  
 CC  
 SQ Sequence 353 AA;  
 Query Match 100.0%; Score 80; DB 16; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 5e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PCVFIRKRVSNVTHG 15  
 DB 106 PCVFIRKRVSNVTHG 120  
 RESULT 6  
 AAR81587  
 ID AAR81587 standard; protein; 353 AA.  
 XX  
 AC AAR81587;  
 XX  
 DT 24-MAY-1996 (first entry)  
 XX  
 DE Cedar pollen allergen B.  
 XX  
 KM Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KM antibody; pollinosis; therapy; immunotherapy.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN EP700929-A2.  
 XX  
 PD 13-MAR-1996.  
 XX  
 PF 08-SEP-1995; 95EP-0306295.  
 XX  
 PR 14-JUL-1995; 95JP-0200221.  
 PR 10-SEP-1994; 94JP-0242137.  
 PR 14-JUL-1995; 95JP-0200204.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI Hino K, Saito S, Taniguchi Y;  
 XX  
 DR WPI, 1996-140976/15.  
 XX  
 PT New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PT used for treating cedar pollinosis  
 XX  
 PS Claim 5; Page 31-32; 36pp; English.  
 XX  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (AAR81586) and B (AAR81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 Peptides (AAR81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (AAR81571-79) essential  
 CC for T-cell recognition, and homologous peptides (AAR81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 CC  
 SQ Sequence 353 AA;  
 Query Match 100.0%; Score 80; DB 17; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 5e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCVFIKRVSNVITIG 15  
 DB 106 PCVFIKRVSNVITIG 120

## RESULT 7

AAV25664 standard; protein; 373 AA.

AAV25664;

30-SEP-1999 (first entry)

Cedar allergen 493634 Cry j IB protein fragment.

Major histocompatibility complex; class II; desensitizing; human;  
 allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 mite; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

Cedrus sp.

WO9934826-A1.

15-JUL-1999.

11-JAN-1999; 99WO-GB00080.

21-SEP-1998; 98GB-0020474.

09-JAN-1998; 98GB-0000445.

(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

Kay AB, Larche M;

WPI; 1999-458255/38.

Desensitizing patients to polypeptide allergens

Example 6; Page 73; 117pp; English.

This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a cedar (Cedrus sp.) allergen 493634 Cry j IB.

Sequence 373 AA;

Query Match 100.0%; Score 80; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 5.4e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 PCVFIKRVSNVITIG 15

RESULT 8

AAV25668 standard; protein; 373 AA.

AAV25668;

30-SEP-1999 (first entry)

Japanese cedar allergen 541803 Cry j I precursor protein fragment.

Major histocompatibility complex; class II; desensitizing; human;  
 allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 mite; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.

Cedrus sp.

WO9934826-A1.

15-JUL-1999.

11-JAN-1999; 99WO-GB00080.

21-SEP-1998; 98GB-0020474.

09-JAN-1998; 98GB-0000445.

(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

Kay AB, Larche M;

WPI; 1999-458255/38.

Desensitizing patients to polypeptide allergens

Example 6; Page 75; 117pp; English.

This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a Japanese cedar (Cedrus sp.) allergen 541803 Cry j I precursor.

Sequence 373 AA;

Query Match 100.0%; Score 80; DB 20; Length 373;

Best Local Similarity 100.0%; Pred. No. 5.4e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 PCVFIKRVSNVITIG 15

126 PCVFIKRVSNVITIG 140

RESULT 9

AAV25668 standard; protein; 374 AA.

AAV25668;

03-JUN-1993 (first entry)

```

XX      ?
DE      Cry j I.
XX      Japanese cedar pollen; allergen; antigen; allergy; B cell; T cell.
XX      Cryptomeria japonica.
OS      Cryptomeria japonica.
XX      Key
FH      Peptide      Location/Qualifiers
FT      1..21
FT      /note= "signal peptide"
FT      Protein      22..374
FT      /note= "mature Cry j I"
XX      MO9301213-A.
XX      21-JAN-1993.
XX      10-JUL-1992; 92WO-US05661.
XX      12-JUL-1991; 91US-0729134.
XX      15-JUL-1991; 91US-0730452.
XX      (IMMU-) IMMUNOLOGIC PHARM CORP.
XX      Bond JF, Griffith IU, Pollock J;
XX      WPI; 1993-045434/05.
XX      N-PSDB; AAQ35304.
XX      Nucleic acid sequence encoding Cryptomeria japonica allergen -
XX      PT for the diagnosis treatment and prevention of allergic reactions
XX      to Japanese cedar pollen
XX      Claim 11; Page 42; 69pp; English.
XX      Fresh pollen and staminate cone samples were collected from a single
XX      CC Cryptomeria japonica (Japanese cedar) tree. RNA was prep'd. and used
XX      CC to synthesize cDNA. The cDNA was subjected to successive rounds of
XX      CC PCR to yield a full length Cry j I clone. Cry j I or an antigenic
XX      CC fragment of it may be used for detecting, treating and preventing an
XX      CC allergic response to Japanese cedar pollen allergen. It is capable of
XX      CC modifying both the B and T cell response to Cry j I and T cell response
XX      CC to a Cry j I antigen.
XX      SQ Sequence 374 AA;
Query Match      100.0%; Score 80; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 5.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 PCVFIRKRVSNVITIG 15
DB      127 PCVFIRKRVSNVITIG 141
RESULT 10
AAR45541
ID      AAR45541 standard; Protein; 374 AA.
XX      AAR45541;
XX      13-JUL-1994 (first entry)
XX      Cry j I pollen allergen.
XX      Japanese cedar; detection; allergy; treatment; diagnosis;
XX      KW T cell epitope; sensitivity.
XX      Cryptomeria japonica.
XX      Key
FH      Peptide      Location/Qualifiers
FT      1..21
FT      /note= "signal peptide"

```

```

FT      Peptide      22..374
FT      /note= "mature peptide"
XX      MO9401560-A.
XX      20-JAN-1994.
XX      15-JAN-1993; 93WO-US00139.
XX      01-SEP-1992; 92US-0938990.
XX      10-JUL-1992; 92WO-US05661.
XX      (IMMU-) IMMUNOLOGIC PHARM CORP.
XX      Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;
XX      WPI; 1994-035066/04.
XX      N-PSDB; AAQ35271.
XX      Antigens derived from Japanese cedar pollen allergen Cry j I -
XX      PT contain at least two T cell epitope(s), used to treat or diagnose
XX      PT allergy
XX      PS Disclosure; Fig 4; 137pp; English.
XX      The sequence is that of the Japanese cedar pollen allergen
XX      CC Cry j I which contains at least two T cell epitopes. Peptide
XX      CC antigens derived from it can be used for the treatment and
XX      CC diagnosis of allergies associated with Japanese cedar pollen.
XX      CC The peptides have enhanced therapeutic properties but reduced
XX      CC side effects compared to naturally occurring allergens.
XX      SQ Sequence 374 AA;
Query Match      100.0%; Score 80; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 5.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 PCVFIRKRVSNVITIG 15
DB      127 PCVFIRKRVSNVITIG 141
RESULT 11
AAR60166
ID      AAR60166 standard; Protein; 374 AA.
XX      AAR60166;
XX      24-MAR-1995 (first entry)
XX      Japanese cedar pollen antigen CryjI.
XX      Japanese cedar; pollen antigen; allergen; CryjI; sugi; pollinosis.
XX      Cryptomeria japonica.
XX      Key
FH      Peptide      Location/Qualifiers
FT      1..21
FT      /label= signal_peptide
FT      Protein      22..374
FT      /label= mature_CryjI
XX      JP06197768-A.
XX      19-JUL-1994.
XX      07-JAN-1993; 93JP-0001116.
XX      07-JAN-1993; 93JP-0001116.
XX      (MEIJ) MEIJI SEIKA KAISHA.

```

DR WPI, 1994-268680/33.  
 DR N-PSDB; AAQ71601.  
 XX  
 PT Sugi (Japanese cedar) pollen antigen Cryj1 - is useful for  
 diagnosis, treatment and prevention of sugi pollinosis  
 XX  
 PS Claim 2; Page 5-7; 9pp; Japanese.  
 XX  
 CC The coding sequence for the Japanese cedar ("sugi") pollen allergen  
 CC Cryj1 was isolated from a cDNA library prepared from polyA mRNA. All  
 CC or part of the Cryj1 protein can be used for diagnosis, treatment  
 CC and prevention of sugi pollinosis.  
 CC  
 SQ Sequence 374 AA;  
 Query Match 100.0%; Score 80; DB 15; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PCVFIKRVSNVTHG 15  
 DB 127 PCVFIKRVSNVTHG 141  
 RESULT 12  
 AAR82490  
 ID AAR82490 standard; Protein; 374 AA.  
 XX  
 AC AAR82490;  
 XX  
 DT 15-APR-1996 (first entry)  
 XX  
 DE Cry j I Japanese Cedar pollen allergen.  
 XX  
 KW Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 KW allergy; Crypomeria japonica.  
 XX  
 OS Crypomeria japonica.  
 XX  
 PN WO9527786-A1.  
 XX  
 PD 19-OCT-1995.  
 XX  
 PE 06-APR-1995; 95MO-US04249.  
 XX  
 PR 06-DEC-1994; 94US-0350225.  
 PR 08-APR-1994; 94US-0226248.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
 PI Shaked Z;  
 XX  
 DR WPI; 1995-366391/47.  
 DR N-PSDB; AAT04248.  
 XX  
 PT Modified Crypomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 XX  
 PS Disclosure; Figure 1; 60pp; English.  
 XX  
 CC Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,  
 CC modified and unmodified, are given in AAR82491-R82525.  
 CC  
 SQ Sequence 374 AA;

Query Match 100.0%; Score 80; DB 16; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PCVFIKRVSNVTHG 15  
 DB 127 PCVFIKRVSNVTHG 141  
 RESULT 13  
 AAY25665  
 ID AAY25665 standard; protein; 374 AA.  
 XX  
 AC AAY25665;  
 XX  
 DT 30-SEP-1999 (first entry)  
 XX  
 DE Cedar allergen 493632 Cry j IA protein fragment.  
 XX  
 KW Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.  
 XX  
 OS Cedrus sp.  
 XX  
 PN WO9934826-A1.  
 XX  
 PD 15-JUL-1999.  
 XX  
 PE 11-JAN-1999; 99WO-GB00080.  
 XX  
 PR 21-SEP-1998; 98GB-0020474.  
 PR 09-JAN-1998; 98GB-0000445.  
 XX  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX  
 PI Kay AB, Larche M;  
 XX  
 DR WPI, 1999-458255/38.  
 XX  
 PT Desensitizing patients to polypeptide allergens  
 XX  
 PS Example 6; Page 73; 117pp; English.  
 XX  
 CC This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents a cedar (Cedrus sp.) allergen 493632 Cry j IA.  
 CC  
 SQ Sequence 374 AA;  
 Query Match 100.0%; Score 80; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PCVFIKRVSNVTHG 15  
 DB 127 PCVFIKRVSNVTHG 141



```

RESULT 14
ID AAY25669
XX AAY25669 standard; protein; 374 AA.
XX
XX AAY25669;
XX
XX 30-SEP-1999 (first entry)
XX
XX Japanese cedar allergen 541802 Cry j I precursor protein fragment.
XX
XX Major histocompatibility complex; class II; desensitizing; human;
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; string;
XX chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
XX screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
XX cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
XX mice; gerbil; vaccine; treatment; prevention; hypersensitivity; cedar.
XX
XX Cedrus sp.
XX
XX WO9334826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GH00080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 75; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
XX polypeptide allergen and comprises administering to the patient a peptide
XX derived from the allergen where restriction to a MHC Class II molecule
XX possessed by the patient can be demonstrated for the peptide and the
XX peptide is able to induce a late phase response in an individual who
XX possesses the MHC Class II molecule. The methods can be used for
XX desensitizing patients to allergens present in e.g. grass, tree and weed
XX (including ragweed) pollens, fungi and moulds, foods, stinging insects,
XX the chironomidae (non-biting midges), spiders and mites, housefly, fruit
XX fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
XX non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
XX tenbrito molitor beetle, mammals such as cat, dog, horse, cow, pig,
XX sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
XX produce immunological vaccines, which may be used to prevent and/or treat
XX conditions involving hypersensitivity to allergens. This sequence
XX represents a Japanese cedar (Cedrus sp.) allergen 541802 Cry j I
XX precursor.
XX
XX Sequence 374 AA;
XX
XX Query Match 100.0%; Score: 80; DB 20; Length 374;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 PCVFIRKRVNVIHG 15
XX |||||
XX 127 PCVFIRKRVNVIHG 141
XX
XX RESULT 15
XX AAM57750
XX ID AAM57750 standard; peptide; 15 AA.
XX

```

```

AC AAM57750;
XX
XX 17-SEP-1998 (first entry)
XX
XX Residues 106-120 of Cry j 1.
XX
XX Cry j 1; Japanese cedar pollen antigen; allergy; immunotherapy;
XX HLA class II molecule.
XX
XX Cryptomeria japonica.
XX
XX WO9820902-A1.
XX
XX 22-MAY-1998.
XX
XX 12-NOV-1997; 97WO-JP04129.
XX
XX 13-NOV-1996; 96JP-0302053.
XX
XX (MEIP ) MEIJI MILK PROD CO LTD.
XX
XX Dairiki K, Kino K, Kume A, Some T;
XX
XX WPI; 1998-297617/26.
XX
XX Peptides derived from Japanese cedar pollen antigens are
XX immunotherapeutic agents - useful for allergy treatment and typing
XX HLA class II molecules in allergy sufferers
XX
XX Claim 12; Page 26; 50pp; Japanese.
XX
XX This sequence represents residues 106-120 of the Cry j 1 protein, and
XX is a peptide of the invention. The peptides are derived from Japanese
XX cedar pollen antigens, and are used as immunotherapeutic agents in the
XX treatment of allergy. The peptides can be used for identification and
XX typing of the particular HLA class II molecules in an allergy sufferer,
XX and also for peptide immunotherapy of an allergy. Using these peptides
XX the immunotherapy can be targeted more specifically to the requirements
XX of the individual patient, allowing more effective treatment of an
XX allergy, including those patients for whom treatment with a conventional
XX immunotherapeutic agent is ineffective.
XX
XX Sequence 15 AA;
XX
XX Query Match 91.3%; Score 73; DB 19; Length 15;
XX Best Local Similarity 93.3%; Pred. No. 2.7e-06;
XX Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 PCVFIRKRVNVIHG 15
XX |||||
XX 1 PCVFIRKRVNVIHG 15
XX
XX Search completed: April 20, 2003, 13:06:08
XX Job time : 19.1974 secs
XX

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 12:54:05; Search time 18.1974 Seconds  
(without alignments)  
109.838 Million cell updates/sec

Title: US-09-142-524d-57

Perfect score: 78

Sequence: 1 KSMKVTVAENQFGPN 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	15	16	AA89293 Japanese cedar pol
2	78	100.0	15	16	AA89293 Residues 211-225 o
3	78	100.0	17	19	AA80349 Sugi allergen prot
4	78	100.0	20	15	AA84563 Cry j I pollen all
5	78	100.0	20	16	AA82512 Cry j I Japanese C
6	78	100.0	20	19	AA84214 T-cell epitope pep
7	78	100.0	26	15	AA845591 Cry j I pollen all
8	78	100.0	26	15	AA845592 Cry j I pollen all
9	78	100.0	26	15	AA845594 Cry j I pollen all
10	78	100.0	28	15	AA845589 Cry j I pollen all

11	78	100.0	28	15	AA845590 Cry j I pollen all
12	78	100.0	28	15	AA845593 Cry j I pollen all
13	78	100.0	30	15	AA845587 Cry j I pollen all
14	78	100.0	30	15	AA845588 Cry j I pollen all
15	78	100.0	30	19	AA84687 T-cell epitope pep
16	78	100.0	36	15	AA845586 Cry j I pollen all
17	78	100.0	353	17	AA875388 Japanese cedar pol
18	78	100.0	353	16	AA881587 Cedar pollen aller
19	78	100.0	354	17	AA804344 Chamaecyparis obtu
20	78	100.0	354	19	AA842121 Japanese cyprus p
21	78	100.0	367	15	AA845577 Jun s I, Juniperu
22	78	100.0	373	20	AA825664 Cedar allergen 493
23	78	100.0	373	20	AA825668 Japanese cedar all
24	78	100.0	374	14	AA831937 Cry j I, Crytome
25	78	100.0	374	15	AA845541 Cry j I pollen all
26	78	100.0	374	15	AA860166 Japanese cedar pol
27	78	100.0	374	16	AA882490 Cry j I Japanese C
28	78	100.0	374	20	AA825665 Cedar allergen 493
29	78	100.0	374	20	AA825669 Japanese cedar all
30	78	100.0	375	17	AA804345 Chamaecyparis obtu
31	73	93.6	38	22	AA884103 Immunomodulatory p
32	73	93.6	74	21	AA823905 Artificial sequenc
33	73	93.6	94	21	AA823906 Artificial sequenc
34	73	93.6	95	21	AA823897 Artificial sequenc
35	73	93.6	95	21	AA823901 Cedar pollen aller
36	73	93.6	95	22	AA869098 Cedar pollen aller
37	73	93.6	96	22	AA869104 Cedar pollen aller
38	73	93.6	96	22	AA869119 Cedar pollen aller
39	73	93.6	97	22	AA869105 Cedar pollen aller
40	73	93.6	99	21	AA823902 Artificial sequenc
41	73	93.6	99	21	AA823907 Artificial sequenc
42	73	93.6	99	22	AA869099 Cedar pollen aller
43	73	93.6	214	22	AA869120 Cedar pollen aller
44	69	88.5	15	22	AA884113 Amino acid sequenc
45	69	88.5	80	18	AA827369 Multi-epitope pept

#### ALIGNMENTS

RESULT 1  
AA89293  
ID AA89293 standard; peptide; 15 AA.  
XX  
XX  
AC AA89293;  
XX  
XX  
12-MAR-1996 (first entry)  
XX  
XX  
Japanese cedar pollen allergen Cryj I derived T-cell epitope peptide.  
DE  
XX  
XX  
Japanese cedar; pollen allergen; Cryj I; T-cell epitope; peptides;  
KW prevention; treatment; cryptomeria pollinosis.  
XX  
XX  
Cryptomeria japonica.  
OS  
XX  
XX  
JP07118295-A.  
PN  
XX  
XX  
09-MAY-1995.  
PD  
XX  
XX  
20-OCT-1993; 93JP-0262626.  
PF  
XX  
XX  
20-OCT-1993; 93JP-0262626.  
PR  
XX  
XX  
20-OCT-1993; 93JP-0262626.  
PA  
XX  
XX  
(MEIP) MEIJI MILK PROD CO LTD.  
XX  
XX  
WPI; 1995-203834/27.  
XX  
XX  
New cryptomeria pollen allergen T-cell epitope peptide - used for  
PT prevention, treatment and investigation of Japanese cedar pollinosis  
XX  
XX  
Claim 5; Page 2; 8pp; Japanese.  
PS  
XX  
XX  
AA875388 is the Japanese cedar pollen allergen Cryj I, from which the

CC T-cell epitope peptides AAR89289-R89295 were derived. The peptides  
 CC can be used for the prevention and treatment of Cryptomeria  
 CC pollenosis, and also for the investigation of pollenosis.

XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 78; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15

RESULT 2  
 ID AAW57755  
 XX AAW57755 standard; peptide; 15 AA.

AC AAW57755;

DT 17-SEP-1998 (first entry)

DE Residues 211-225 of Cry j 1.

KM Cry j 1; Japanese cedar pollen antigen; allergy; immunotherapy;  
 KM HLA class II molecule.

OS Cryptomeria japonica.

PN WO9820902-A1.

PD 22-MAY-1998.

PF 12-NOV-1997; 97WO-JP04129.

PR 13-NOV-1996; 96JP-0302053.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Kuno K, Kume A, Sone T;

DR WPI; 1998-297617/26.

PT Peptides derived from Japanese cedar pollen antigens are

PT immunotherapeutic agents - useful for allergy treatment and typing

PS HLA class II molecules in allergy sufferers

PS Claim 12; Page 28; 50pp; Japanese.

CC This sequence represents residues 211-225 of the Cry j 1 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.

SQ Sequence 15 AA;

Query Match 100.0%; Score 78; DB 19; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.2e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15

RESULT 3

AAW80349  
 ID AAW80349 standard; peptide; 17 AA.

AC AAW80349;

DT 11-JAN-1999 (first entry)

DE Sugi allergen protein Cryj2 derived epitope for T cells.

KM T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;

KM sugi-pollenosis; allergic reaction; pollen.

OS Synthetic.

PN JP10259198-A.

PD 29-SEP-1998.

PF 22-DEC-1997; 97JP-0353448.

PR 24-DEC-1996; 96JP-0343441.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PA (SANY) SANKYO CO LTD.

DR WPI; 1998-577037/49.

PT A linked T cell epitope peptide - used for the treatment of

PT sugi-pollenosis

PS Claim 7; Page 18; 21pp; Japanese.

CC AAW80339-58 represent epitopes for T cells, derived from the sugi  
 CC allergen proteins Cryj1 (AAW80339-44, AAW80350-53 and AAW80356-58) and  
 CC Cryj2 (AAW80345-49 and AAW80354-55). The peptides are useful for the  
 CC treatment of sugi-pollenosis, an allergic reaction of the body to  
 CC pollen.

SQ Sequence 17 AA;

Query Match 100.0%; Score 78; DB 19; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
 DB 2 KSMKVTVAFNQFGPN 16

RESULT 4

ID AAR45563 standard; Protein; 20 AA.

AC AAR45563;

DT 13-JUL-1994 (first entry)

DE Cry j 1 pollen allergen peptide CJI-22.

KM Japanese cedar; detection; allergy; treatment; diagnosis;

KM T cell epitope; sensitivity.

OS Cryptomeria japonica.

PN WO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 XX WPI; 1994-035066/04.  
 DR  
 XX  
 PT Antigens derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 PS Claim 1; Fig 13; 137bp; English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I (amino acids 211-230). The peptide, CJI-22,  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens.  
 CC  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 78; DB 15; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15  
 RESULT 5  
 ID AAR82512 standard; Protein; 20 AA.  
 AC AAR82512;  
 XX  
 DT 15-APR-1996 (first entry)  
 XX  
 DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-22).  
 XX  
 XX Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 KM allergy; Crytpomeria japonica.  
 KM  
 OS Crytpomeria japonica.  
 XX  
 PN WO9527786-A1.  
 XX  
 PD 19-OCT-1995.  
 XX  
 PF 06-APR-1995; 95WO-US04249.  
 XX  
 PR 06-DEC-1994; 94US-0350225.  
 PR 08-APR-1994; 94US-0226248.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
 PI Shaked Z;  
 XX  
 DR WPI; 1995-366391/47.  
 PT Modified Crytpomeria japonica (Cry j) I peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 XX  
 PS Disclosure; Figure 2; 60pp; English.  
 CC Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j I,

CC modified and unmodified, are given in AAR82491-R82525. This peptide  
 CC fragment corresponds to amino acids 211-230 of the allergen mature  
 CC protein.  
 CC  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 78; DB 16; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15  
 RESULT 6  
 ID AAW42144 standard; peptide; 20 AA.  
 AC AAW42144;  
 XX  
 DT 16-JUN-1998 (first entry)  
 XX  
 DE T-cell epitope peptide 24 from Japanese cypress pollen antigen Chaol.  
 XX  
 XX Japanese cypress pollen; antigen; T-cell epitope; Chaol; Chaol2;  
 KW diagnosis; allergy; spring tree pollen disease; pollinosis.  
 XX  
 OS Chamaeyparis obtusa.  
 OS  
 PN WO9747648-A1.  
 XX  
 PD 18-DEC-1997.  
 XX  
 PF 12-JUN-1997; 97WO-JP02031.  
 XX  
 PR 14-JUN-1996; 96JP-0153527.  
 XX  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 XX  
 PI Dairiki K, Kino K;  
 PI  
 DR WPI; 1998-052242/05.  
 XX  
 PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
 PT Chaol and Chaol2 - used for diagnosis and treatment of spring tree  
 PT pollen disease  
 XX  
 PS Claim 1; Page 27-28; 71pp; Japanese.  
 CC The present sequence represents a T-cell epitope peptide from Japanese  
 CC cypress pollen antigen Chaol. The present invention describes peptides  
 CC which correspond to the T-cell epitope sites on Japanese cypress pollen  
 CC antigens Chaol and Chaol2. The peptides can be used as a reagent for the  
 CC diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
 CC the treatment and prevention of spring tree pollen disease in which the  
 CC pollinosis involves reactivity to Japanese cypress pollen.  
 CC  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 78; DB 19; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15  
 RESULT 7  
 ID AAR45591 standard; Protein; 26 AA.  
 XX

AC AAR45591;  
 XX  
 DT 13-JUL-1994 (first entry)  
 XX  
 DE Cry j I pollen allergen peptide CJI-43.9.  
 XX  
 KM Japanese cedar; detection; allergy; treatment; diagnosis;  
 XX T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 XX  
 PN WO9401560-A.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US00139.  
 XX  
 PR 01-SEP-1992; 92US-0938990.  
 XX 10-JUL-1992; 92WO-US05661.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;  
 XX WPI; 1994-035066/04.  
 DR  
 XX Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 PS Claim 76; Fig 18; 137pp; English.  
 CC  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I. The peptide, CJI-43.9, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 XX  
 SQ Sequence 26 AA;  
 Query Match 100.0%; Score 78; DB 15; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15  
 RESULT 8  
 AAR45592  
 ID AAR45592 standard; Protein; 26 AA.  
 XX  
 AC AAR45592;  
 XX  
 DT 13-JUL-1994 (first entry)  
 XX  
 DE Cry j I pollen allergen peptide CJI-43.10.  
 XX  
 KM Japanese cedar; detection; allergy; treatment; diagnosis;  
 XX T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 XX  
 PN WO9401560-A.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US00139.  
 XX  
 PR 01-SEP-1992; 92US-0938990.  
 XX 10-JUL-1992; 92WO-US05661.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;  
 XX WPI; 1994-035066/04.  
 DR  
 XX Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 PS Claim 76; Fig 18; 137pp; English.  
 CC  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I. The peptide, CJI-43.10, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 XX  
 SQ Sequence 26 AA;

PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 DT Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;  
 XX WPI; 1994-035066/04.  
 DR  
 XX Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 PS Claim 76; Fig 18; 137pp; English.  
 CC  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I. The peptide, CJI-43.10, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 XX  
 SQ Sequence 26 AA;  
 Query Match 100.0%; Score 78; DB 15; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KSMKVTVAFNQFGPN 15  
 DB 1 KSMKVTVAFNQFGPN 15  
 RESULT 9  
 AAR45594  
 ID AAR45594 standard; Protein; 26 AA.  
 XX  
 AC AAR45594;  
 XX  
 DT 13-JUL-1994 (first entry)  
 XX  
 DE Cry j I pollen allergen peptide CJI-43.12.  
 XX  
 KM Japanese cedar; detection; allergy; treatment; diagnosis;  
 XX T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 XX  
 PN WO9401560-A.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US00139.  
 XX  
 PR 01-SEP-1992; 92US-0938990.  
 XX 10-JUL-1992; 92WO-US05661.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;  
 XX WPI; 1994-035066/04.  
 DR  
 XX Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 XX  
 PS Claim 76; Fig 18; 137pp; English.  
 CC  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j I. The peptide, CJI-43.12, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 XX  
 SQ Sequence 26 AA;

Query Match 100.0%; Score 78; DB 15; Length 26;  
Best Local Similarity 100.0%; Pred. No. 7.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
DB 1 KSMKVTVAFNQFGPN 15

## RESULT 10

AAR45589 standard; Protein; 28 AA.

AC AAR45589;

DT 13-JUL-1994 (first entry)

DE Cry j I pollen allergen peptide CUI-43.7.

KW Japanese cedar; detection; allergy; treatment; diagnosis;  
KM T cell epitope; sensitivity.

OS Crytomeria japonica.

PN WO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;

DR WPI; 1994-035066/04.

PT Antigen derived from Japanese cedar pollen allergen Cry j I -  
PT contain at least two T cell epitope(s), used to treat or diagnose  
PT allergy

PS Claim 76; Fig 18; 137pp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j I. The peptide, CUI-43.7, can be used for  
CC the treatment and diagnosis of allergies associated with Japanese  
CC cedar pollen. It has enhanced therapeutic properties but reduced  
CC side effects compared to naturally occurring allergens.

SQ Sequence 28 AA;

Query Match 100.0%; Score 78; DB 15; Length 28;

Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
DB 1 KSMKVTVAFNQFGPN 15

## RESULT 11

AAR45590 standard; Protein; 28 AA.

AC AAR45590;

DT 13-JUL-1994 (first entry)

DE Cry j I pollen allergen peptide CUI-43.8.

KW Japanese cedar; detection; allergy; treatment; diagnosis;

KW T cell epitope; sensitivity.

OS Crytomeria japonica.

PN WO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;

DR WPI; 1994-035066/04.

PT Antigen derived from Japanese cedar pollen allergen Cry j I -  
PT contain at least two T cell epitope(s), used to treat or diagnose  
PT allergy

PS Claim 76; Fig 18; 137pp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j I. The peptide, CUI-43.8, can be used for  
CC the treatment and diagnosis of allergies associated with Japanese  
CC cedar pollen. It has enhanced therapeutic properties but reduced  
CC side effects compared to naturally occurring allergens.

SQ Sequence 28 AA;

Query Match 100.0%; Score 78; DB 15; Length 28;

Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQFGPN 15  
DB 1 KSMKVTVAFNQFGPN 15

## RESULT 12

AAR45593 standard; Protein; 28 AA.

AC AAR45593;

DT 13-JUL-1994 (first entry)

DE Cry j I pollen allergen peptide CUI-43.11.

KW Japanese cedar; detection; allergy; treatment; diagnosis;

KM T cell epitope; sensitivity.

OS Crytomeria japonica.

PN WO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;

DR WPI; 1994-035066/04.

PT Antigen derived from Japanese cedar pollen allergen Cry j I -

PT contain at least two T cell epitope(s), used to treat or diagnose  
allergy

PS Claim 76; Fig 18; 137bp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j I. The peptide, CJI-43.11, can be used for  
CC the treatment and diagnosis of allergies associated with Japanese  
CC cedar pollen. It has enhanced therapeutic properties but reduced  
CC side effects compared to naturally occurring allergens.

XX Sequence 28 AA;

SQ

Query Match 100.0%; Score 78; DB 15; Length 28;  
Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
DB 1 KSMKVTVAFNQFGPN 15

RESULT 13

AA45587  
ID AAR45587 standard; Protein; 30 AA.

AC AAR45587;

DT 13-JUL-1994 (first entry)

DE Cry j I pollen allergen peptide CJI-43.1.

XX Japanese cedar; detection; allergy; treatment; diagnosis;

KW T cell epitope; sensitivity.

OS Cryptomeria japonica.

XX WO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

PA Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;

PI WPI, 1994-035066/04.

DR WPI, 1994-035066/04.

XX Antigen derived from Japanese cedar pollen allergen Cry j I -

PT contain at least two T cell epitope(s), used to treat or diagnose

PT allergy

PS Claim 76; Fig 18; 137bp; English.

XX The sequence is that of an isolated peptide of the Japanese cedar

CC pollen allergen Cry j I. The peptide, CJI-43.1, can be used for

CC the treatment and diagnosis of allergies associated with Japanese

CC cedar pollen. It has enhanced therapeutic properties but reduced

CC side effects compared to naturally occurring allergens.

XX Sequence 30 AA;

SQ

Query Match 100.0%; Score 78; DB 15; Length 30;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15  
DB 1 KSMKVTVAFNQFGPN 15

RESULT 14

AA45588  
ID AAR45588 standard; Protein; 30 AA.

AC AAR45588;

DT 13-JUL-1994 (first entry)

DE Cry j I pollen allergen peptide CJI-43.6.

XX Japanese cedar; detection; allergy; treatment; diagnosis;

KW T cell epitope; sensitivity.

OS Cryptomeria japonica.

XX WO9401560-A.

PD 20-JAN-1994.

PF 15-JAN-1993; 93WO-US00139.

PR 01-SEP-1992; 92US-0938990.

PR 10-JUL-1992; 92WO-US05661.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

PA Bond JF, Garman RD, Griffith IU, Kuo M, Pollock J;

PI WPI, 1994-035066/04.

DR WPI, 1994-035066/04.

XX Antigen derived from Japanese cedar pollen allergen Cry j I -

PT contain at least two T cell epitope(s), used to treat or diagnose

PT allergy

PS Claim 76; Fig 18; 137bp; English.

XX The sequence is that of an isolated peptide of the Japanese cedar

CC pollen allergen Cry j I. The peptide, CJI-43.6, can be used for

CC the treatment and diagnosis of allergies associated with Japanese

CC cedar pollen. It has enhanced therapeutic properties but reduced

CC side effects compared to naturally occurring allergens.

XX Sequence 30 AA;

SQ

Query Match 100.0%; Score 78; DB 15; Length 30;

Best Local Similarity 100.0%; Pred. No. 9.2e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSMKVTVAFNQFGPN 15

DB 1 KSMKVTVAFNQFGPN 15

RESULT 15

AA44687  
ID AAW44687 standard; peptide; 30 AA.

AC AAW44687;

DT 01-MAY-1998 (first entry)

DE T-cell epitope peptide #6 of sugi pollen antigen.

XX T-cell epitope; sugi pollen antigen; sugi pollinosis.

XX Synthetic.

OS Cryptomeria japonica.

XX JP10007700-A.

PN 13-JAN-1998.



XX 24-JUN-1996; 96JP-0163287.  
 PF  
 XX 24-JUN-1996; 96JP-0163287.  
 PR  
 XX (DAIL ) DAICEL CHEM IND LTD.  
 PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
 XX  
 DR WPI; 1998-133630/13.  
 XX  
 PT T cell epitope peptide of sugi pollen antigen - useful in the  
 PT treatment of sugi pollinosis  
 XX  
 PS Claim 1; Page 4; 14pp; Japanese.  
 XX  
 CC T-cell epitope peptides AAW4682-88 and their derivatives react with  
 CC sugi pollinosis patient peripheral blood T lymphocytes. A composition  
 CC prepared by combining at least 2 of the above peptides and/or their  
 CC derivatives is used for the prevention and treatment of sugi  
 CC pollinosis.  
 XX  
 SQ Sequence 30 AA;

Query Match 100.0%; Score 78; DB 19; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSMKVTVAFNQGFN 15  
 |||||  
 DB 11 KSMKVTVAFNQGFN 25

Search completed: April 20, 2003, 13:06:08  
 Job time : 18.1974 secs

